Naval Research Laboratory
Stennis Space Center, MS 39529-5004

NRL/FR/7323-92-9412

Circulation Model Experiments of the Gulf Stream Using Satellite-Derived Fields

DANIEL N. FOX
MICHAEL R. CARNES
JIMMY L. MITCHELL

Ocean Dynamics and Prediction Branch
Oceanography Division

April 1993

REPORT DOCUMENTATION PAGE

Form Approved
OBM No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1993	Final

4. TITLE AND SUBTITLE

Circulation Model Experiments of the Gulf Stream Using Satellite-Derived Fields

5. FUNDING NUMBERS

Job Order No.	73509103
Program Element No.	0602435N
Project No.	RM35G93
Task No.	201
Accession No.	DN251050

6. AUTHOR(S)

Daniel N. Fox, Michael R. Carnes, and Jimmy L. Mitchell

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Oceanography Division
Stennis Space Center, MS 39529-5004

8. PERFORMING ORGANIZATION REPORT NUMBER

Formal Report 9412
(NOARL Report 55)

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Code 124
800 N. Quincy Street
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The ability to perform skillful forecasts of the space and time evolution of mesoscale ocean phenomena can significantly improve acoustic surveillance capabilities, search and rescue operations, the performance of weapon systems, and other aspects of naval operations. The Ocean Dynamics and Prediction Branch, Ocean Monitoring and Prediction Section of the Naval Research Laboratory (NRL), Stennis Space Center, MS, is developing an ocean nowcast/forecast capability to permit dynamical interpolation of the limited number of in situ and satellite observations available using a hierarchy of coupled hydrodynamic and thermodynamic models.

A series of numerical experiments to predict the evolution of the Gulf Stream were performed. These experiments led to the first system to show significant skill in this area. The forecast system uses the NRL Northwest Atlantic regional primitive equation model and assimilation schemes. The schemes employ both a feature model and statistical correlations derived from the regional climatology of in situ data and long time-base numerical simulations. The evaluation criterion is the mean absolute distance between forecast locations of the Gulf Stream front and actual locations, as verified from extensive satellite and in situ data. Eight 1-week and five 2-week evaluation intervals during 1986–1988 were selected to represent a variety of both active and inactive Gulf Stream regimes. To ensure objectivity, hindcasting was disallowed. The NRL system provided forecasts that were significantly better than persistence at both 1- and 2-week intervals. This study confirms the feasibility of Gulf Stream forecasting using assimilation schemes that provide adequate deep information and numerical models that are designed to be consistent with available data.

14. SUBJECT TERMS

ocean models, ocean forecasting, fronts (oceanography), air-sea interaction

15. NUMBER OF PAGES

46

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Same as report

CONTENTS

CIRCULATION MODEL EXPERIMENTS OF THE GULF STREAM USING SATELLITE-DERIVED FIELDS

1.0 INTRODUCTION

The western boundary current regions of the oceans represent domains of high variability and considerable eddy activity. They are not adequately sampled by in situ measurements or by one single-beam satellite altimeter (Hurlburt 1986; Kindle 1986) to permit accurate, instantaneous estimates of sea surface topography. They are also regions of strategic importance, both economically and militarily. The development of numerical models of the North Atlantic (Thompson and Schmitz 1989) capable of reproducing the measured variabilities, large-scale circulations, and eddy activity has given us the confidence to proceed with the construction of a system that will use such models to dynamically interpolate the asynoptic measurements available. This capability will thereby provide improved nowcast and forecast capabilities.

The Data Assimilation Research and Transition (DART) Project at the Naval Research Laboratory (NRL), developed a modular end-to-end nowcast/forecast system. This system maximizes the use of existing operational Navy analysis and forecast modules. When fully implemented, the system outlined in Fig. 1 will continuously assimilate satellite and in situ data (including expendable bathythermographs, or XBTs, and temperatures inferred from acoustic tomography) into a set of coupled models that define the circulation, the gross thermal structure, and the fine mixed layer structure of the ocean. All the models—respectively known as the Ocean Circulation: Evolution, Assimilation and Nowcasting System (OCEANS); the Optimum Thermal Interpolation System (OTIS); and the Thermodynamic Ocean Prediction System (TOPS)—were developed at Navy laboratories and operational centers. All have become standard in several, as-yet uncoupled, Navy operational modules. The DART Project focuses its efforts on linking these modules into a viable nowcast/forecast system that will operate in both the central site and shipboard environments.

To provide a baseline against which to measure future progress, a preliminary Gulf Stream version of the system (hereafter referred to as OCEANS/GS, version 1.0) was constructed. The forecast skill of this system was evaluated using synoptic datasets that define the evolution of the Gulf Stream front and rings during a series of five 2-week reference periods beginning in late 1986 and ending in mid-1988. These reference datasets were jointly developed by NRL and Harvard University for an evaluation of the Harvard GulfCast system (Robinson et al. 1987) by Commander, Naval Oceanography Command (COMNAVOCEANCOM). The evaluation was done in mid-1989 and was presented to the COMNAVOCEANCOM Interim Model Review and Evaluation Panel (CIMREP) in December 1989. Figures 2 and 3 and Table 1 summarize the results of this evaluation. It was found that the DART Gulf Stream system provided a significant forecast enhancement in the Gulf Stream frontal location over persistence (the assumption of no change) for at least 2 weeks (the extent of the evaluation periods), but the GulfCast system did not.

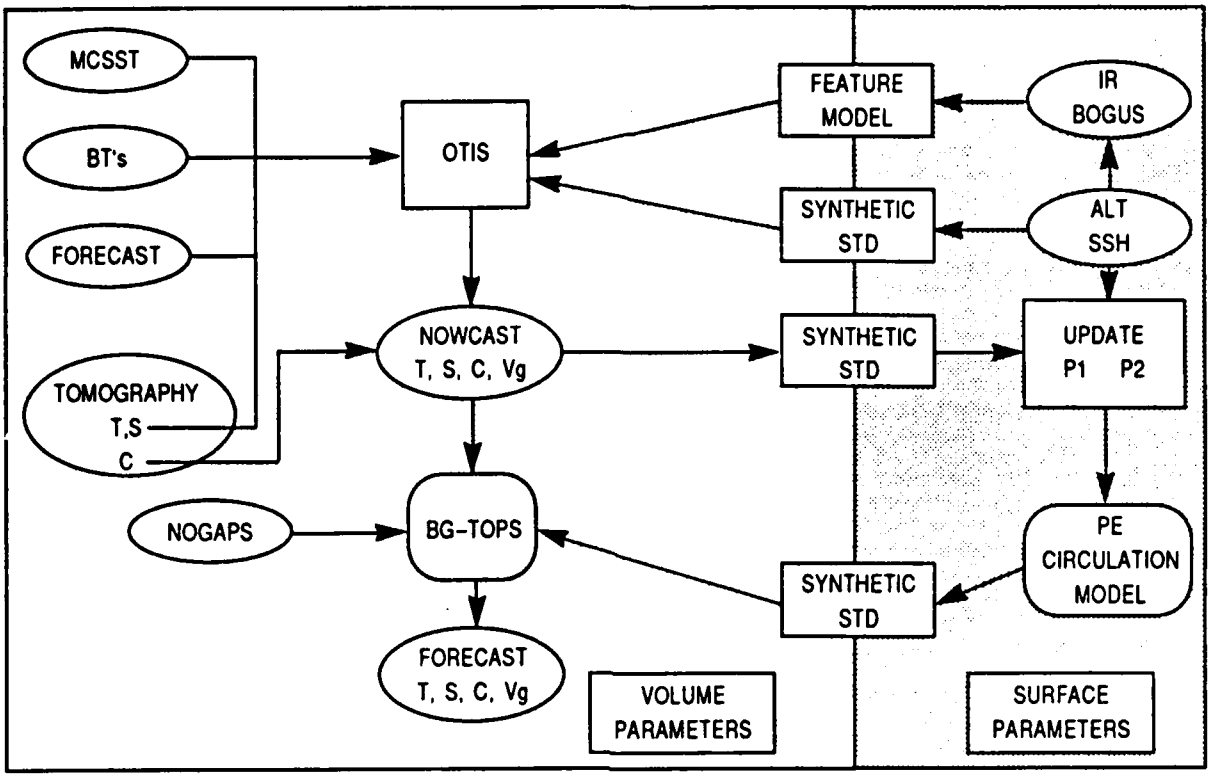

Fig. 1 — Schematic of the complete *DART OCEANS/GS nowcast/forecast* system. See Fig. 4 for a diagram of the subset of this system which was evaluated in this report.

Figure 4 highlights the components of the OCEANS/GS 1.0 system used in this evaluation. XBT and multichannel sea surface temperature (MCSST) data were used to provide front and eddy information. The OTIS feature model software was then used to convert this information into a three-dimensional thermal volume, which is integrated into a dynamic height at the surface. This height field is used to initialize a circulation model forecast. Details on this procedure are given in the following sections.

Considerable effort went to examining the sensitivity to errors in the initial state of the DART forecast. Section 4.3 describes the procedure whereby the accuracy of the initial states was estimated. Each initial state was then modified to yield a series of several alternate but equally valid representations of the positions of the axis and eddies. Forecasts were made with each of these initial states to isolate regions where forecasts were (and were not) sensitive to errors in the initial state. The results of these experiments suggest the value of a Monte Carlo approach to performing forecasts.

Section 2 summarizes the results of the evaluation of OCEANS/GS 1.0. It compares the skill of this system against persistence and the previous version of the Navy Operational Gulf Stream Forecast System (NOGUFS) (Rhodes and Horton 1990) product. Section 3 gives details on the construction of the initial states and the mechanics of performing the forecasts. Section 4 examines the accuracy of the initial states and the effect of the errors on the forecasts. Last, Section 5 discusses computer resource issues.

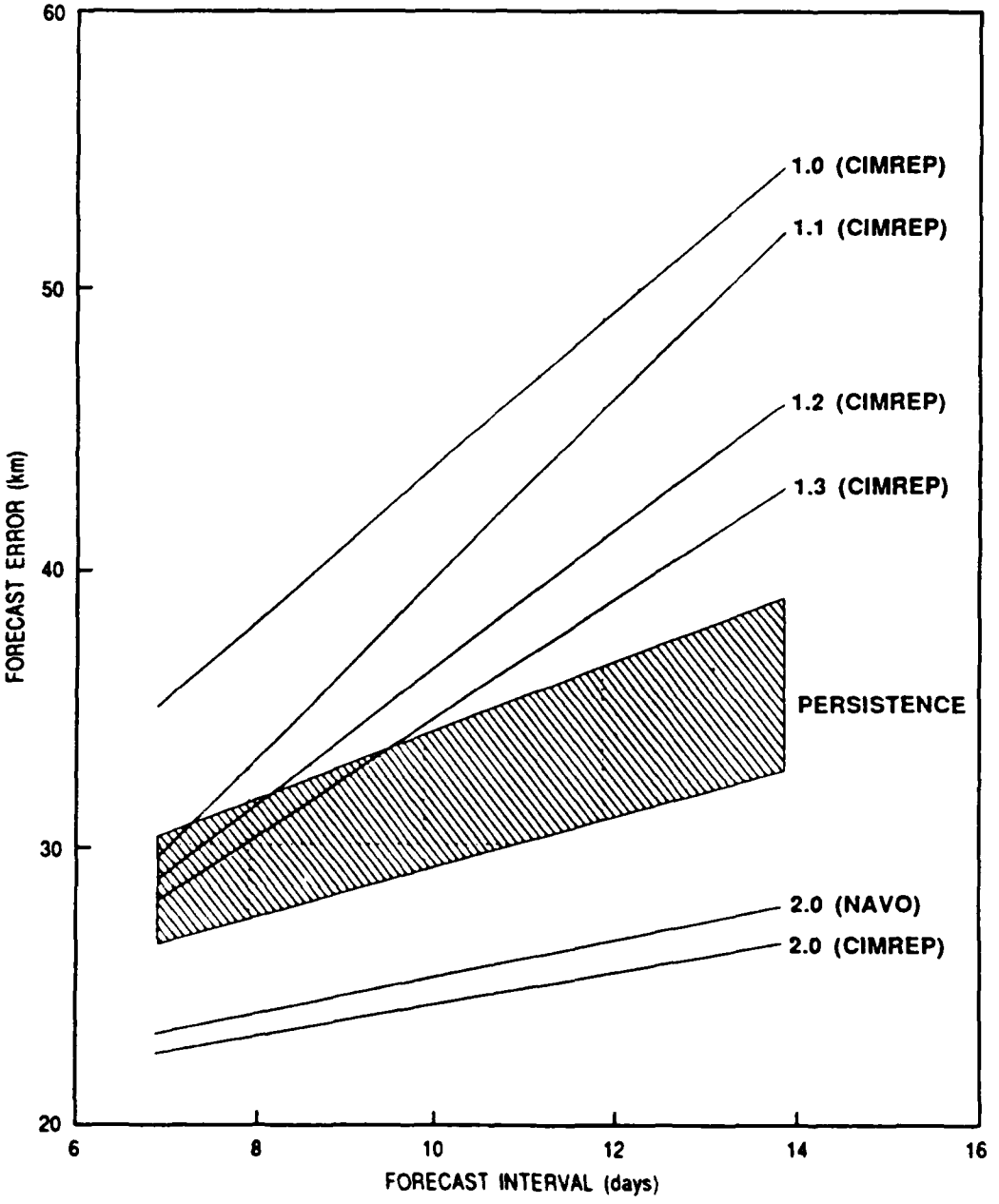

Fig. 2 — NOGUFS forecast skill compared to persistence. Version 2.0 is the DART system described in this report. Versions 1.0 through 1.3 represent earlier versions which were based on the Harvard GulfCast system. Forecast skill using the research grade datasets are labeled "CIMREP." When applied to a set of operational data (labeled "NAVO"), the DART system continued to show skill relative to persistence.

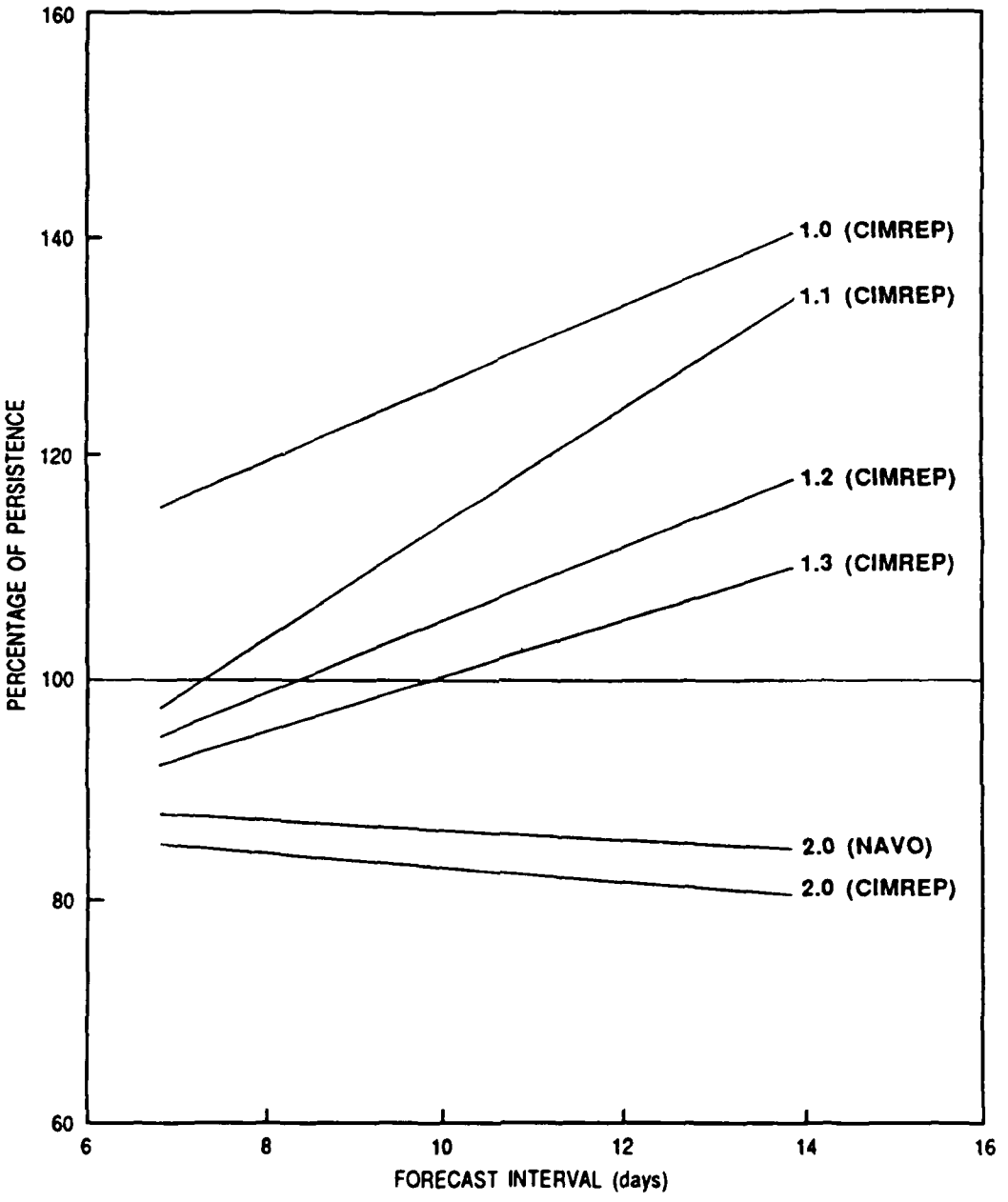

Fig. 3 — NOGUFS forecast skill expressed as a percentage of persistence (showing the range of operational enhancement), comparing the earlier versions (1.x, based on GulfCast) to 2.0 (DART). See Fig. 2 caption for details.

Table 1 — Statistical Significance of Forecast Results

Analysis of Forecasts								
	DART OCEANS/GS 1.0				NOGUFS 1.3 (GulfCast)			
Period	MEAN	Std Err	P (type I)	P > 0	MEAN	Std Err	P (type I)	P > 0
1 Week	4.5 km	2.2 km	0.007	0.978	1.9 km	3.4 km	0.742	0.705
2 Weeks	8.1 km	3.7 km	0.125	0.987	−7.1 km	6.2 km	0.375	0.128

Analysis of the degree to which each model beats persistence for 1- and 2-week forecasts over the region from 73°W to 53°W. To show significant skill, the mean must be positive and the P(type 1) value must be small. See Sec. 2.3 for discussion.

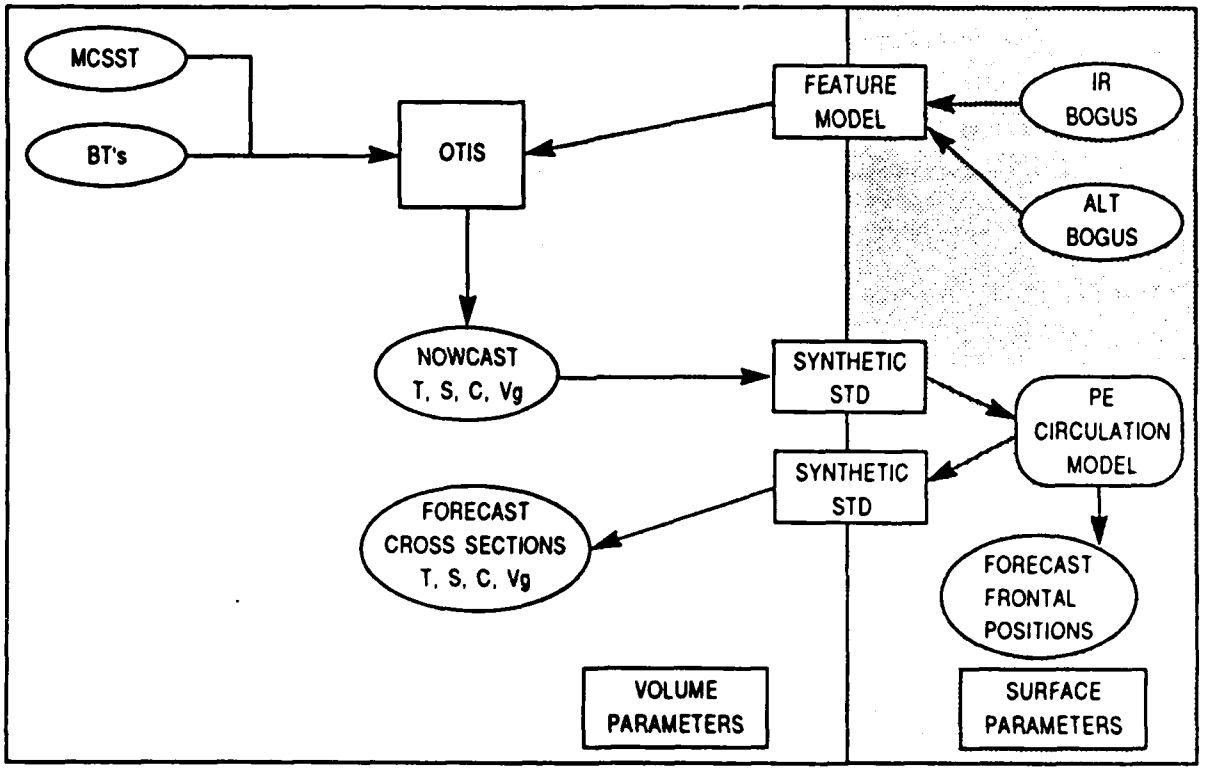

Fig. 4 — Schematic of the version of the DART OCEANS/GS forecast system as evaluated in this report.

This report is not an evaluation of the Harvard GulfCast system. But because NOGUFS 1.0 was based on that model, comparisons are both inevitable and appropriate. For details on the Navy's evaluations of GulfCast, the reader is referred to Rhodes and Heburn (1990) and to Rhodes and Horton (1990).

Since the version of the GulfCast model as originally delivered by Harvard was unable to provide forecasts that were better than persistence when applied to operational datasets, NRL, the Naval Oceanographic Office (NAVOCEANO), and Harvard University jointly made several changes in an attempt to improve its forecast skill, denoted by versions 1.0 through 1.3 of NOGUFS

(Rhodes and Heburn 1990; Rhodes and Horton 1990). Version 1.3, referred to as the "9-level, intermediate, extended domain" model, nominally showed the greatest forecast skill of the four versions, so it was chosen for the comparisons made in this report.

2.0 RESULTS OF EVALUATION

In this section, we summarize the results of the evaluation of the DART OCEANS/GS 1.0 in comparison to persistence and to the previous version of NOGUFS, which was based on Harvard's GulfCast system. We present these results prior to a complete description of the DART OCEANS/GS 1.0 system (see Sec. 3).

2.1 Forecast Experiments

To provide a "level playing field" in the arena of ocean forecasting, a set of reference data (i.e., Gulf Stream frontal locations) was created for a subset of the Gulf Stream region. Before any forecasts had been done, representatives from NRL, Harvard University, and a contractor third party[1] examined the existing data over the preceding 3 years to define several periods of at least 2 weeks during which accurate positions of the Gulf Stream axis and eddies could be constructed. Clear satellite infrared imagery was required and, in most cases, XBTs and Geosat altimetry (Born et al. 1987) was used to refine the locations of the features. Table 2 lists the eight 1-week and five 2-week forecast periods extracted from this dataset.

The primary evaluation criterion was the mean absolute distance between forecast locations of the Gulf Stream front and the actual (or verification) locations. Forecast error was computed as the

Table 2 — Dates Used to Initialize and to Verify Forecasts

1-Week Forecasts		2-Week Forecasts	
Initial State	Verification	Initial State	Verification
11/26/86	12/03/86	11/26/86	12/10/86
04/08/87	04/15/87	04/08/87	04/22/87
04/15/87	04/22/87	04/22/87	05/06/87
05/06/87	05/13/87	05/06/87	05/20/87
05/13/87	05/20/87		
07/08/87	07/15/87	07/08/87	07/22/87
07/15/87	07/22/87		
05/04/88	05/11/88		

[1] R. Crout, Planning Systems, Inc., Slidell, LA.

average absolute offset between the forecasted position of the axis with the position given in the verification state. (Details of this methodology are given in Sec. 2.2) *Persistence*, the assumption of no change over the forecast interval, was used as a comparison reference in judging forecast skill. For a model to have any significant skill in forecasting, the error in its forecast (the *forecast error*) must be less than the error obtained by using the initial state as the forecast (the *persistence error*).

To examine the degree to which this small number of states was representative of the Gulf Stream as a whole, persistence errors were computed both from these states and from a 1-year series of weekly frontal positions prepared at the Naval Eastern Oceanography Center (NEOC). Figures 5 and 6 show (among other things) histograms of the computed persistence errors for 1- and 2-week delays and verify (by comparison with an annual time series of NEOC frontal axis

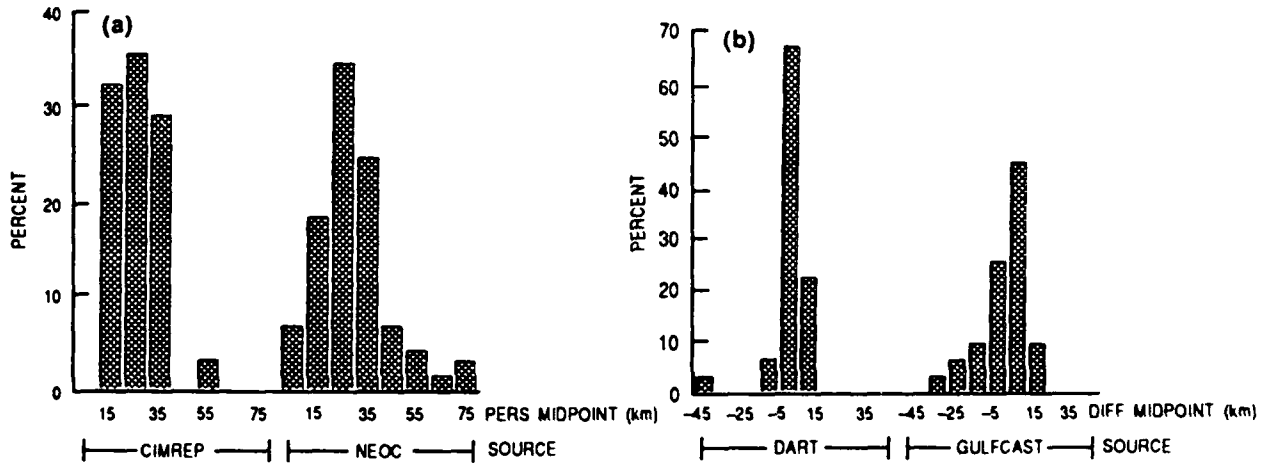

Fig. 5 — (a) Histogram comparing persistence errors in the reference datasets to those computed from a year of NEOC front locations. (b) Comparison of the distribution of forecast error from the DART system and NOGUFS. This figure uses data from the 1-week forecast periods.

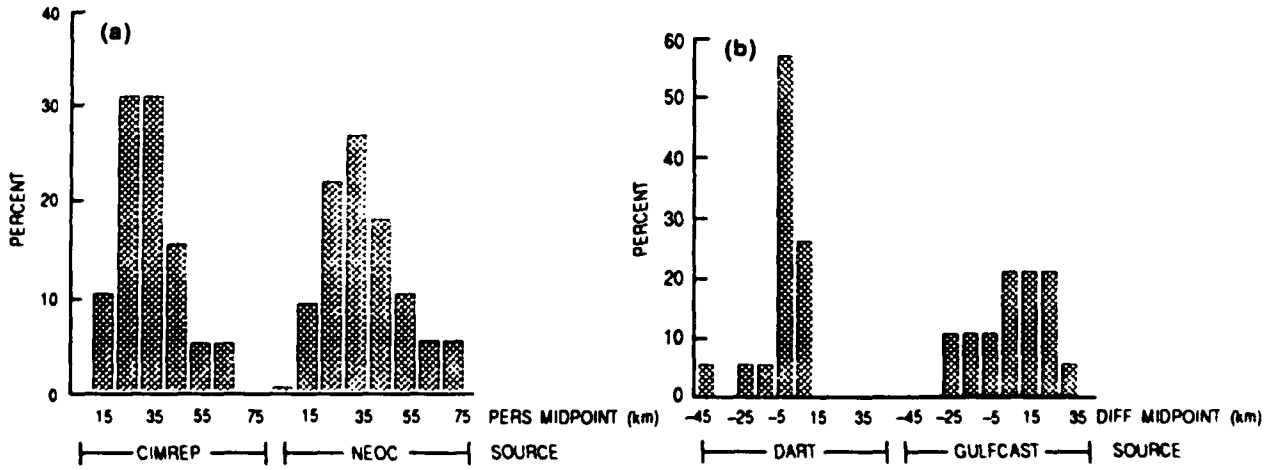

Fig. 6 — (a) Histogram comparing persistence errors in the reference datasets to those computed from a year of NEOC front locations. (b) Comparison of the distribution of forecast error from the DART system and NOGUFS. This figure uses data from the 2-week forecast periods.

maps) that the reference evaluation states probably are representative of the Gulf Stream. The distributions of persistence error for the NEOC and the special CIMREP evaluation cases can be compared using standard statistical methods (Crow et al. 1960). Table 3 shows the 95% confidence range for the difference between the means of the two populations. The only region in which this range does not include zero is in the east, where the stream is the most variable. Even in this region, however, reducing the confidence level to 90% will result in a range that includes zero. We can thus be fairly confident that the CIMREP reference states are typical of the Gulf Stream.

It should be noted that the NEOC boguses were the routine operational product; they generally did not include as much information as was available for the evaluation reference dates.

2.2 Verification Methodology

As mentioned in the previous section, the forecast skill of the model was quantified by computing the average absolute offset between the position of the Gulf Stream axis in the model forecast and in the verification state for that day.

Table 3 — Analysis of Differences Between Distributions of Persistence Error Computed from the NEOC Boguses and the CIMREP Cases, Showing that the CIMREP Cases are Representative

1-Week Persistence Errors (km)					
	CIMREP Cases ($N = 8$)		NEOC Cases ($N = 49$)		Difference (95% Confidence)
Region	$\bar{x}_C$	σ_C	$\bar{x}_N$	σ_N	$\bar{x}_C - \bar{x}_N$
West	22.7	11.9	26.3	14.4	-3.7 ± 10.5
Center	29.0	12.7	26.5	10.4	2.5 ± 8.1
East	35.7	21.4	26.9	7.3	8.8 ± 7.8
Overall	29.4	10.9	27.2	5.5	2.3 ± 4.9
2-Week Persistence Errors (km)					
	CIMREP Cases ($N = 5$)		NEOC Cases ($N = 48$)		Difference (95% Confidence)
Region	$\bar{x}_C$	σ_C	$\bar{x}_N$	σ_N	$\bar{x}_C - \bar{x}_N$
West	30.3	15.3	32.9	17.5	-2.6 ± 15.9
Center	37.8	14.1	30.4	13.4	7.4 ± 12.3
East	48.7	22.9	35.5	7.6	13.2 ± 9.3
Overall	39.4	11.3	33.3	10.3	6.1 ± 9.6

This offset was chosen as the prime evaluation criterion rather than pattern correlation because of the nature of the initial and verification states. Figure 7 shows a sample initial state derived by a method that uses simple models of the Gulf Stream and eddies (feature models) and that tends to be somewhat cartoon-like or schematic in character. Figure 8, by comparison, shows a field from a model simulation run. Many small recirculation features present in the forecast (and in the real ocean) are not present in the feature model initial and verification states, which would yield low pattern correlations even if the forecast were exactly predicting the motion of the axis and the eddies. Further, since the limited area model of the Gulf Stream does not include the *basin-scale* North Atlantic ocean recirculation, the motion of the eddies was not expected to properly forecast either. Therefore, the evaluations described here do not address the issue of isolated ring propagation.

These considerations lead us to adopt the same evaluation criterion used by the CIMREP panel to measure the performance of the GulfCast system: that is, the average absolute offset of the predicted axis location from its "true" location (as given by the verification state).

The offset error is computed by calculating the *area* (in square kilometers) bounded by the two axes being compared, then dividing by the *length* of the axis taken as the truth. Figure 9 shows an example of this process. The solid line represents the axis taken from the verification state, and the dotted line is the forecast starting from the previous week. The lower panel shows the shaded area that the program computes. These computations were performed using software created by Harvard University (Gardner 1989) and modified by the authors for use with the DART forecast system.

The model domain was divided into three subregions for the evaluations: a western region extending from 73°W to 66°W, a central region extending from 66°W to 59°W, and an eastern region extending from 59°W to 53°W. The average absolute offset error was computed for each of these subregions, as well as for the overall region, which extended from 73°W to 53°W.

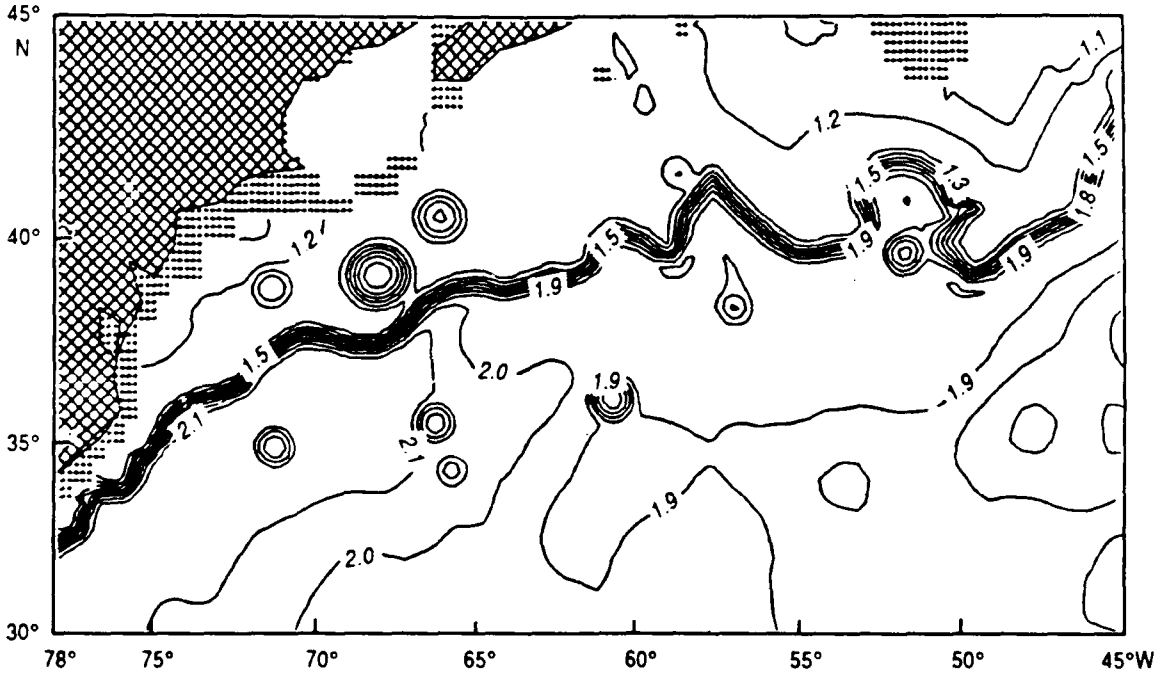

Fig. 7 — Feature-modeled surface topography from a NEOC front and eddy map for 15 April 1987

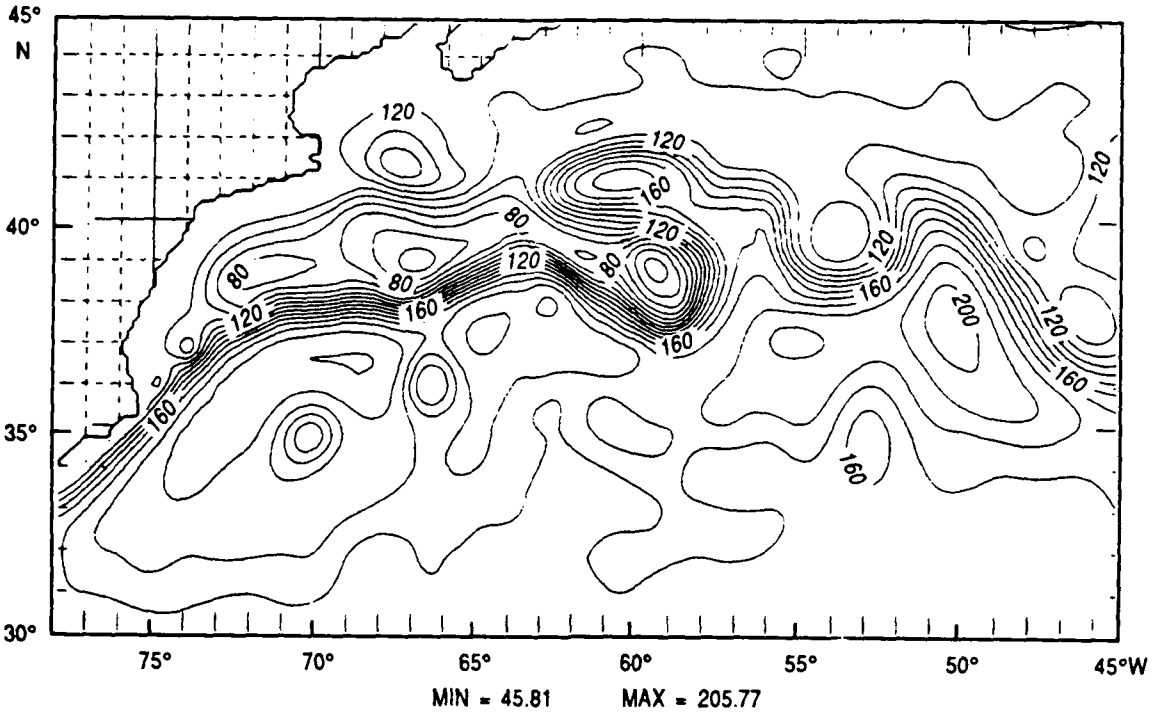

Fig. 8 — Plot of a sample sea surface height field from a circulation model forecast. A bias of 1.5 m has been added to convert the field into an approximation of dynamic height for comparison with Figs. 7 and 10.

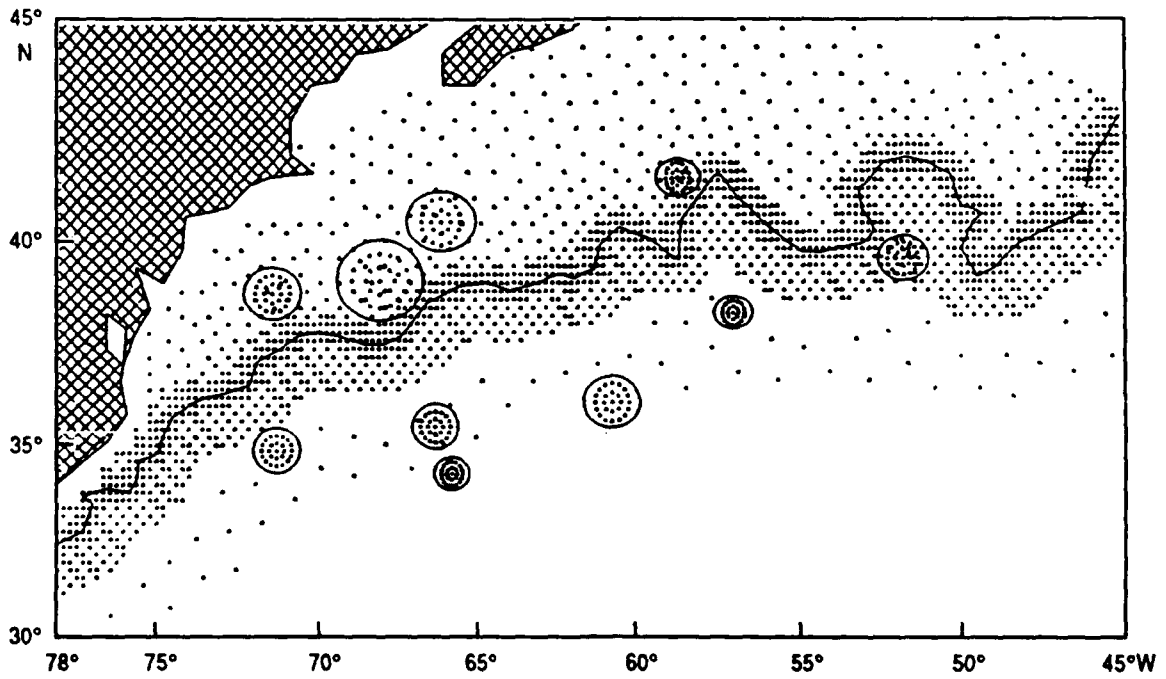

Fig. 9 — The positions of the synthetic points generated by OTIS to generate the field shown in Fig. 7.

2.3 Results

Table 4 shows the computed average absolute offset errors for the DART OCEANS/GS 1.0 for each of the eight 1-week and five 2-week forecast periods. Each pair of numbers represents the forecast error (the error between the forecast position of the axis and that given by the verification state), followed by the persistence error (the error computed if the initial state is used unchanged as the forecast). For comparison, Table 5 shows the same information computed using NOGUFS version 1.3.

Table 4 — Raw Data from DART Forecasts on Reference States

DART OCEANS/GS 1.0									
Forecast and Persistence Errors at +1 Week (km)									
		Region							
Initial	Forecast	Overall		West		Center		East	
		FCST	PERS	FCST	PERS	FCST	PERS	FCST	PERS
861126	861203	26.7	27.2	16.9	16.2	36.0	37.8	-na-	-na-
870408	870415	22.4	22.6	23.0	18.5	10.6	16.5	31.1	30.7
870415	870422	19.5	20.4	29.0	32.2	13.1	14.4	16.1	14.8
870506	870513	20.0	38.6	13.5	58.3	18.3	19.5	29.6	35.7
870513	870520	25.0	28.7	12.7	19.0	31.2	35.0	29.8	30.1
870708	870715	25.0	26.0	30.4	29.9	19.2	18.1	25.8	29.9
870715	870722	21.5	24.2	19.0	22.5	30.2	30.5	15.6	19.0
880504	880511	21.6	29.7	10.6	14.2	25.2	39.9	27.8	27.7
Forecast and Persistence Errors at +2 Weeks (km)									
		Region							
Initial	Forecast	Overall		West		Center		East	
		FCST	PERS	FCST	PERS	FCST	PERS	FCST	PERS
861126	861210	28.7	32.9	21.4	19.7	35.0	44.8	-na-	-na-
870408	870422	22.5	27.9	16.6	24.0	16.1	24.5	33.1	34.0
870422	870506	24.9	24.2	28.7	25.6	17.6	12.7	24.8	30.8
870506	870520	30.4	50.7	17.4	63.1	40.5	42.5	30.5	46.6
870708	870722	27.6	30.5	23.3	31.9	31.2	27.4	26.7	30.5

Table 5 — Raw Data from NOGUFS (GulfCast) 1.3 Forecasts on Reference States

NOGUFS 1.3 (GulfCast)									
Forecast and Persistence Errors at +1 Week (km)									
		Region							
Initial	Forecast	Overall		West		Center		East	
		FCST	PERS	FCST	PERS	FCST	PERS	FCST	PERS
861126	861203	40.0	32.5	20.7	17.8	58.5	46.5	-na-	-na-
870408	870415	34.1	25.6	38.2	20.7	23.2	19.0	38.8	33.5
870415	870422	28.0	23.5	30.1	35.4	21.1	17.6	31.2	17.9
870506	870513	22.8	40.5	28.2	52.5	24.5	20.9	13.4	47.5
870513	870520	23.9	31.9	11.6	24.0	28.1	35.7	30.8	33.4
870708	870715	25.3	28.5	40.1	33.4	8.7	17.6	29.2	35.0
870715	870722	31.4	26.3	27.6	26.8	43.3	35.2	23.9	16.8
880504	880511	23.1	34.5	12.7	17.6	20.3	44.0	41.2	33.8
Forecast and Persistence Errors at +2 Weeks (km)									
		Region							
Initial	Forecast	Overall		West		Center		East	
		FCST	PERS	FCST	PERS	FCST	PERS	FCST	PERS
861126	861210	38.4	38.4	44.8	21.4	31.4	53.8	-na-	-na-
870408	870422	43.0	32.1	26.0	25.3	46.1	28.0	54.5	40.6
870422	870506	54.2	26.0	51.3	24.6	51.2	17.3	56.8	33.6
870506	870520	43.4	52.1	37.1	60.8	48.6	43.0	41.7	54.2
870708	870722	40.3	35.6	25.8	35.9	50.6	36.8	41.6	31.9

The quantity being analyzed in each case is the difference between the error in the model forecast and the error in the persistence forecast. Positive differences indicate that the model forecast is better than persistence and that negative differences occur when the model forecast is worse than persistence. Since the two models integrate different variables, the stream axis is defined differently: the zero-line of the stream function in the uppermost level in the case of the GulfCast quasi-geostrophic circulation model, the zero contour of the field of upper layer pressure anomaly in the case of the DART primitive equation circulation model. Since each model defines the axis slightly

differently, consistency was maintained by comparing each model forecast to its own estimate of persistence.

Both the initial and verification states contained errors, an issue addressed in Sec. 4; however, the forecast skill is computed with the assumption that the axis positions are known exactly. Since the initial states were prepared by a third party who had no foreknowledge of how the stream should evolve, there is no reason to believe that these initial states should favor one model over the other. In fact, the initial states for these particular tests use stream *axis* location, which the GulfCast feature model accepts directly.

The DART system, however, was designed around operational Navy products. One of these modules, the OTIS feature model, requires the *north wall* of the stream rather than the axis. The transformation from axis to north wall was done simply by translating the axis a fixed number of kilometers normal to its original position. If anything, this procedure caused somewhat of a disadvantage to the DART forecast results but only for these specially prepared test cases. The operationally produced front and eddy maps (such as those prepared by NEOC and NAVOCEANO) provide the north wall directly, which eliminates the need for this translation and thereby eliminates a possible source of error in the proposed operational version of DART.

Table 1 presents a summary of the forecast skill of the DART OCEANS/GS 1.0 and of version 1.3 of NOGUFS. Given the relatively small number of standard cases used to evaluate the models, a matter of particular relevance is the degree of confidence or statistical significance to attach to the results. Table 1 addresses this issue directly and will be described more fully in the following paragraphs; in summary, however, the table shows that while the DART OCEANS/GS 1.0 model forecast error is significantly less than the persistence forecast error for both 1- and 2-week forecasts, the NOGUFS 1.3 model forecast error is not significantly less than the persistence forecast error at 1 week and is actually significantly greater than persistence error for 2-week forecasts.

To attach true statistical significance to the results, the raw data presented in Tables 4 and 5 were analyzed using standard statistical methods, the SAS procedure UNIVARIATE (SAS Institute 1988). Since the number of experiments was small, the distributions could not confidently be assumed to be Gaussian, so nonparametric estimates (based on the actual distribution of the data being analyzed) were used.

The "Mean" column in Table 1 is the difference between the error in the model forecast and the error in the persistence forecast. A positive number indicates that the model error is less than persistence error by that number of kilometers. A negative number indicates that the model error is greater than persistence error. The "Std Err" column is an estimate of the standard error of that mean. The third and fourth columns are estimates of the significance of the results. The third column, P(type I), is the probability of making a Type I error: that is, the probability that in the given small sample, the computed difference between the forecast and persistence error is nonzero, when in fact there is no difference. The fourth column, P>0, is the probability that the mean is positive: that is, the average probability that the model forecast will beat persistence. It is computed from the mean and the standard error of the mean and from the assumption that at least the estimate of the mean is normally distributed. It is the integrated area under the distribution of the mean from zero to positive infinity. A value of 100% here would mean that the model would certainly beat persistence on the average. A value of 0% would mean that the model would certainly

NOT beat persistence on the average, and a value of 50% would mean a toss-up, with the errors in the model forecast and persistence being the same.

As an example, in the 1-week forecasts, the NOGUFS 1.3 model beats persistence by 1.9 km, but the standard error of this mean is 3.4 km, indicating that the result is probably not significant. The probability that the mean is positive (and therefore that the model beats persistence on the average) is 71%, but remember that the 50% level is where the model and persistence become a toss-up. Using the original data distribution, the probability of making a Type I error is 74%, indicating that the null hypothesis (that the model is not significantly different from persistence) cannot be ruled out. The DART model for this same period beats persistence by 4.5 km, and the standard error of this mean is only 2.2 km, which results in a 98% probability (on the average) that the model will beat persistence and a <1% chance of making a Type I error. Thus, even though both models nominally beat persistence at 1 week on the average, only the DART results are statistically significant.

Figures 5 and 6 present an alternate way to visualize the skill of the forecast systems in comparison to persistence. Along with verifying that the reference states are representative of the Gulf Stream system (see Sec. 2.1), they also provide a visual confirmation of the facts presented in Table 1. In particular, the relatively wide, scattered distribution of the forecast errors for the GulfCast model and the relatively narrow distribution of the DART forecast errors clearly show why the DART forecast results achieve statistical significance, but the GulfCast results do not. Even though GulfCast nominally beats persistence at 1 week on the average, the distribution of forecast errors is so large that this result is not significant.

In summary, Figs. 5 and 6 and Table 1 explicitly show that the GulfCast model does not beat persistence in a statistically significant way at 1 week, and loses to persistence in a significant way at 2 weeks. The DART OCEANS/GS 1.0 model beats persistence in a statistically significant way at both 1 and 2 weeks.

3.0 DART GULF STREAM NOWCAST/FORECAST SYSTEM

In this section, we provide a detailed description of the DART OCEANS/GS run stream as diagrammed in Fig. 4. The procedure used to initialize and run the DART OCEANS/GS forecasts begins with the subjective preparation of an initial Gulf Stream frontal location map (see discussions in Sec. 3.1). This manually prepared map blends frontal location information contained in satellite infrared imagery, satellite altimetry, and any available bathythermographs into a continuous depiction of the surface frontal location. These maps are the common starting line for GulfCast, DART OCEANS/GS and persistence runs. For the DART OCEANS/GS, an offset is applied to these surface frontal locations to approximate north wall positions (see Sec. 3.1). The resulting continuous depiction of the north wall location is then run through the OTIS 2.1 Feature Model software. The software provides an initial state estimate of the dynamic height, and the estimate is then converted directly to an upper layer pressure anomaly (p_1) in the NRL primitive equation circulation model. Scaling between OTIS dynamic height and model p_1 is necessary to correctly represent the transport in the model's thick upper layer (see Sec. 3.2). The initial lower layer pressure field (p_2) is derived from the initial p_1 field using a statistical inference technique based upon the circulation model's climatology (see Sec. 3.4). Together, p_1 and p_2 are then used to provide for geostrophic initialization of the circulation model (Sec. 3.5). For a brief interval immediately following these

"cold-start" initializations, a gravity wave filter is applied in the circulation model run (Sec. 3.6). Finally, the $p_1 = 0$ contour in the model's forecast state is used to define the forecast frontal location for direct comparison with independent verification frontal location maps.

Much of the success in these cold-start DART OCEANS/GS forecasts is a result of the several techniques used to directly and instantaneously transfer upper layer information into the lower layer of the circulation model. These techniques are described chronologically in the following sections.

3.1 Feature Model

Surface topography maps are used to initialize the circulation model and to verify the forecasts prepared using the regional OTIS. OTIS was developed primarily at the Fleet Numerical Oceanography Center (FLENUMOCEANCEN), with contributions from NAVOCEANO and NRL (Cummings 1989; Clancy et al. 1988; Bennett and May 1988; Bennett et al. 1988). OTIS is a data quality control and interpolation system that combines climatology, maps of front and eddy boundaries, MCSSTs, and measured temperature profiles to form gridded three-dimensional synoptic thermal analyses for selected ocean regions. A reduced set of the OTIS system capabilities was used in this study: the surface topography maps were produced using maps of front and eddy positions as the only data source. The OTIS software interprets these maps and applies models for the Gulf Stream front and eddies to form a gridded three-dimensional field of temperature. Relative dynamic height at the surface is then computed directly from the grid of temperature profiles using relationships derived from analysis of regional historical temperature and salinity datasets.

The front and eddy maps delineate the path of the Gulf Stream and the radial fringes of rings. These maps were prepared from composites of infrared images, AXBTs, and Geosat altimeter data for each required analysis date. The periods used in this study were selected based upon the availability of large numbers of AXBTs and particularly because of the relatively small amounts of cloud coverage. Images nearest in time to each analysis period were obtained from Channel 4 of the Advanced Very High Resolution Radiometer. The images, which are available four times per day, were displayed on an image processing system. The positions of the surface thermal front boundary were extracted and stored manually by an operator using an interactive cursor. Positions were also extracted from previous images, if possible, when portions of the Gulf Stream or eddies were obscured by clouds on the primary infrared image. Although images could be displayed at the full 1-km resolution, system operators suggest that the accuracy of the frontal position extraction process is about 5 km. Altimeter-derived measurements of surface topography were processed using the NRL Geosat Ocean Applications Program (Lybanon and Crout 1987).

After receiving instrument-error corrected altimeter data from the Applied Physics Laboratory at Johns Hopkins University, further processing is done to remove tides and electromagnetic bias. The initialization and verification dates were chosen during periods when there were little or no cloud cover, and altimetry was used only to precisely locate the Gulf Stream, which has a very strong signal. Consequently, no correction for the effects of atmospheric water vapor was performed. Finally, estimates of the ocean surface topography are computed by removing the geoid height deviation using the high-resolution geoid prepared by NAVOCEANO. However, since the along-track topography profiles are used only to identify positions of mesoscale features, no attempt is made to correct for the long-wavelength orbit height error. The processed height profiles were then

displayed along Geosat ground tracks on the video processing system with a recent infrared image displayed in the background. The edges of the front and of eddies were determined from the height sections and stored digitally.

AXBT data used in this study were collected from four sources: FLENUMOCEANCEN, NEOC, the National Oceanographic Data Center (NODC), and the NRL Regional Energetics Experiment. Each profile was classified by its thermal structure as being located north, south, or within the Gulf Stream, or within a cold- or warm-core ring.

All data for each analysis date were combined onto a single map, which consisted of frontal path segments from infrared, locations of front and ring crossings from altimetry, and AXBT locations coded according to water type. An unbroken frontal path, located from the western to the eastern boundaries of the model domain, was hand-drawn through the composite data set and then digitized, and ring radii and center locations were extracted. Figure 7 is an example of a front and eddy map prepared for 15 April 1987 that shows the position of the surface thermal front of the Gulf Stream and several warm- and cold-core rings.

OTIS produces on the order of 2000 temperature profiles (synthetic profiles) positioned at selected gridpoints throughout the analysis domain. A typical distribution of synthetic profile positions is shown in Fig. 9. Each is derived from models of the Gulf Stream, the rings, and the background water structure using the front and eddy map as a guide. Most synthetic profiles are positioned in the stream, the eddies, and in a band on either side of the stream. The distance between profiles is roughly proportional to the expected temperature covariance length scales at each position.

OTIS uses parametric models of the Gulf Stream front, eddies, and the ambient background. These models were developed from a combination of historical observations, simple dynamical models, and information obtained from published studies. The background temperature field $T^{AMB}(z)$ is prepared, first by modifying the gridded temperature climatology, $T^{CLM}(z)$, with a water-mass-based gridded climatology, $T^{WM}(z)$, such that

$$T_j^{AMB}(z) = T^{CLM}(z) + (T_j^{WM}(z) - T_j^{CLM}(z))\, e^{-(D_j/L)^2} \tag{1}$$

at gridpoint j at position (x_j, y_j), where x is the longitude and y is the latitude at the gridpoint. D_j is the shortest distance from position j to the front; $L = 300$ km is a length scale that controls the blending of the two climatologies; T^{CLM} is the temperature profile from the Generalized Digital Environmental Model (GDEM) (Teague et al. 1990) and from the Navy Standard ocean climatology, interpolated to the gridpoint and to the analysis day; and T^{WM} is a temperature profile from a seasonal water-mass database that contains averages of the most frequently occurring profiles found nearby but outside either side of the front. The effect of this blending is to remove the broad climatological Gulf Stream front from GDEM and replace it at all depths with a step transition, $\Delta T^{WM} = T^{Sargasso} - T^{slope}$, along the path of the north wall. The temperature on each side of the step is then blended smoothly outward into the GDEM climatology. The broad Gulf Stream front of the unaltered GDEM climatology is displayed in the plot of dynamic height (0/2000 dbars) for 15 April (Fig. 10).

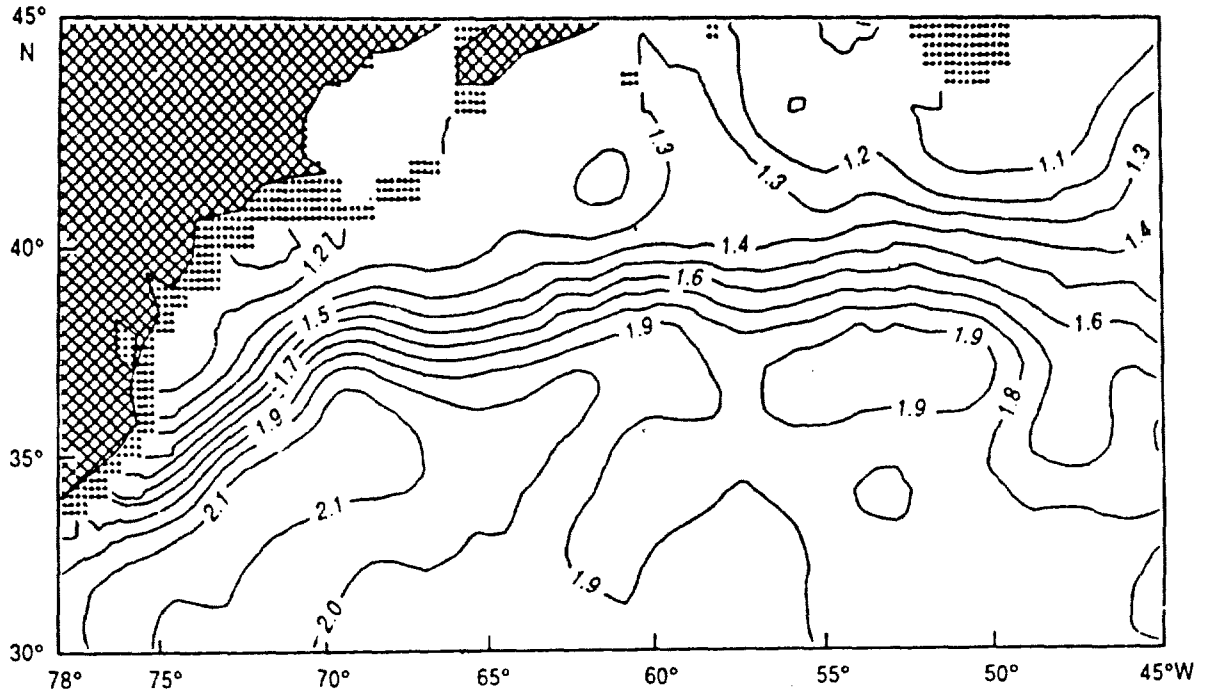

Fig. 10 — Interpolated GDEM climatology for 15 April 1987.

The step transition across the front is replaced using a cross-sectional Gulf Stream front model. This model varies with shortest distance to the north wall, depth, distance downstream along the front, time of year, front curvature, and temperature difference across the step front at each depth. The model front is also designed to approximately conserve potential vorticity as the stream curvature changes along its path.

The Gulf Stream front model is given by

$$T(z) = T^{AMB}(z) - \Delta T^{WM} F(z) + C(z) \tag{2}$$

where T^{AMB} is computed by Eq. (1), T^{WM} is the temperature difference across the step front,

$$F(z) = \begin{cases} e^{-(x_f(z)/W_f(z))^2} & x_f(z) \geq 0 \\ 1 & x_f(z) < 0 \end{cases} \tag{3}$$

produces the smooth temperature transition, and

$$C(z) = (C_{\min} + C_{\max}(1 + e^{(-(P/P_{wc})^2)} \cos((t - 62)\pi/180)))e^{(-(x_f(z)/W_c)^2 - (z/z_c)^2)} \tag{4}$$

inserts the near-surface warm core of the stream.

The term

$$x_f(z) = D_j - x_0(z) \tag{5}$$

is the distance south of the northern extent of the front. The northern edge of the subsurface-modeled front slopes away from the surface boundary location according to

$$x_0(z) = \begin{cases} z/(sz_b) & z \le z_b \\ (1 - (z_e/z_b))(1 - e^{((z_b - z)/z_e)})/s & z_b \le z \le z_b + 25z_e \\ (1 - (z_e/z_b))/s & z \ge z_b + 25z_e \end{cases}, \tag{6}$$

where the depth scales are $z_e = 6z_l$, $z_b = 20z_l$, $z_l = z_{l_1} + z_{l_2} \cos((t - 31)\pi/180)$ with t = day of year, $z_{l_1} = 82.5$ m and $z_{l_2} = 42.5$ m.

The front slope factor is

$$s = 0.02/(1 + 20\lambda), \tag{7}$$

where

$$\lambda = \frac{d\theta}{dp}\Big|_{p = p_j} \tag{8}$$

is the path curvature of the surface front boundary curve at its closest approach to the analysis position (x_j, y_j). θ is the path direction, and p is the distance along the path from where it crosses 27.5°N.

The front width scale in Eq. (4) is defined for each of several depth ranges:

$$W(z) = \begin{cases} 20(1 + 10\lambda) & z < z_{mld} \\ 30(1 + 10\lambda) & z = z_{mld} \\ (22 + z^{0.511})(1 + 10\lambda) & z_{mld} < z < 200\,\text{m} \\ (35 + 0.01z)(1 + 10\lambda) & 200\,\text{m} < z < 1500\,\text{m} \\ (50 + ((z - 1500)/100)^{0.511})(1 + 10\lambda) & 1500\,\text{m} < z < 3000\,\text{m} \\ (51 + 0.001z)(1 + 10\lambda) & 3000\,\text{m} < z < 5000\,\text{m} \end{cases} \tag{9}$$

The Gulf Stream near-surface warm core model, Eq. (4), uses minimum and maximum temperature differences between the ambient water and the warm core defined by $C_{min} = 0.5°$, $C_{max} = 2.5°$. The parameter that controls the weakening of the warm core down stream is $P_{w_l} = 2800$ km, and the width and depth scales for the warm core are $W_c = 53.8$ km, $z_c = 1.8\,z_l$.

The ring model is based on two observations (Joyce 1984): that the surface of the ring rotates as a solid body from the center out to radius R_1, and that the azimuthal velocity decreases linearly with radius from R_1 to R_2. The current-gradient dynamic balance

$$fv + \frac{v^2}{r} = \frac{\partial H}{\partial r}, \tag{10}$$

where f is the Coriolis parameter, v is the azimuthal velocity, r is the radius, and H is the dynamic height, is integrated after substitution of the assumed velocity structure,

$$v = \begin{cases} \omega r & 0 \le r \le R_1 \\ \omega R_1 \dfrac{(r - R_2)}{(R_1 - R_2)} & R_1 \le r \le R_2 \end{cases} \tag{11}$$

to obtain a solution for the dynamic height,

$$H = \begin{cases} H_0 + (f\omega + \omega^2) r^2 & 0 \le r \le R_1 \\ H_{AMB} - (f\omega + \omega^2) R_1^2 - (R_2 - R_1)^2 (fB/2 + B^2/2) + & \\ \qquad (R_2 - R_1)(fR_2 B + 2R_2 B^2) + B^2 R_2^2 \ln (R_1/R_2) & R_1 \le r \le R_2 \\ H_{AMB} & r \ge R_2 \end{cases}, \tag{12}$$

where

$$H_0 = H_{AMB} - (f\omega + \omega^2) R_1^2 - (R_2 - R_1)^2 (fB/2 + B^2/2) + (R_2 - R_1)(fR_2 B + 2R_2 B^2) + B^2 R_2^2 \ln (R_1/R_2) \tag{13}$$

and

$$B = \frac{\omega R_1}{R_1 - R_2}, \tag{14}$$

and ω is the angular velocity of the core, and H_{AMB} is the dynamic height at this position before insertion of the ring. In this study, we set R_1 equal to the surface-temperature-front radius delineated in the front and eddy map and $R_2 = 1.5\ R_1$. The azimuthal velocity is assumed to reach a maximum at R_1 of 1 m/s. Once the ring topography is computed, the temperature structure is derived from a set of regional monthly relationships between dynamic height and vertical structure of temperature. The relationships were derived (deWitt 1987; Carnes et al. 1990) from all available profiles and were truncated to a 1000-m depth in the Navy archive database (NODC hydrographic data is a subset) for the region bounded by 35°N to 40°N by 60°W to 75°W. All profiles, not just those

within rings, were used. To simplify regressions, monthly subsets of profiles were first compressed by representing them with empirical orthogonal functions (EOF) (Davis 1976; Preisendorfer 1988)

$$T(z_k) = \overline{T}(z_k) + \sum_{i=1}^{M} A_i E_i(z_k), k = 1, 19 \tag{15}$$

and truncating to $M = 2$ of the 19 available terms. T is the original or estimated profile, z_k are the 19 standard depths from the surface to 1000 m, $\overline{T}$ is the mean temperature computed over all profiles in the subset, E_i represents the EOFs computed as the eigenvectors of the temperature covariance matrix, and A_i are the EOF amplitudes that distinguish each profile. Truncation of the EOF series to two terms results in approximations to the true temperature profiles, and mean-square error is less than 5%. Least-squares polynomial regression between the amplitudes and dynamic height H was performed to find approximate relations.

$$\hat{A}_i = a_o + \sum_{k=1}^{3} a_k H^k, i = 1,2 \tag{16}$$

The ring profiles were then derived by calculating H from Eq. (12), from which the EOF amplitudes were computed by Eq. (16) and then substituted into Eq. (15).

The OTIS system assimilates synthetic observations, true observations taken at irregular times and positions, and climatology to form synoptic maps using optimum interpolation (Gandin 1963; Bretherton et al. 1976). The synthetic observations provide a high spatial resolution data set within and near fronts and eddies, where observational data are often too sparse to resolve these features. Also, the synthetic profiles provide subsurface information, whereas most measurements are made only at the surface from satellites. This study uses only synthetic profiles to construct the final temperature field. Optimum interpolation estimates the temperature anomaly $\hat{T}'_j$ at position (x_j, y_j) from a weighted sum

$$\hat{T}'_j = \sum_{i=1}^{N} W_{ij} \tilde{T}'_i \tag{17}$$

of observed (or synthetic) temperature anomalies

$$\tilde{T}'_i = \tilde{T}_i - \overline{T}_i \tag{18}$$

at positions (x_i, y_i), where $\overline{T}$ is the expected temperature and $\tilde{T}$ is the measured (or synthetic) temperature. The anomaly of the measured profile

$$\tilde{T}'_i = T'_i + \epsilon_i^I - \epsilon_i^N \tag{19}$$

is composed of the true temperature anomaly plus instrument measurement error ϵ_i^I minus the variability ϵ_i^N at wavenumbers higher than those resolved by the final interpolated grid. The weights

W_{ij} are derived by minimizing the squared error of the estimate $\hat{T}'$ over an ensemble of trials, with the result

$$W_{ij} = \sum_{k=1}^{N} [A^{-1}]_{ik} B_{jk} , \tag{20}$$

where $[A^{-1}]_{ik}$ is the (i,k) component of the inverse of the observation anomaly covariance matrix

$$A_{ik} = \overline{\tilde{T}'_i \tilde{T}'_k} = S_{ij}^2 + \left(I_{ik}^2 + N_{ik}^2\right)\delta_{ik} \tag{21}$$

and B_{jk} is the covariance between the true anomaly values of temperature at the analysis position j and the kth observation anomaly,

$$B_{jk} = \overline{T'_j \tilde{T}'_k} = S_{jk}^2 . \tag{22}$$

For the mean value, $\overline{T}$, we use the climatological profile from GDEM interpolated to the position and the day of year. The instrument error variance is fixed at $I^2 = 1°C^2$ for all synthetic profiles. Both the noise variance and signal covariance are fixed a priori and change with location, but they are held constant within a given province. The separate provinces are the regions north of the Gulf Stream, south of the stream, the front, and cold- and warm-core rings. The signal covariance between two positions within the same province is defined to have a Gaussian form and may be anisotropic.

$$S_{ij}^2 = \sigma_k^2 e^{-(\Delta X_{ij}/R_{Xk})^2 - (\Delta Y_{ij}/R_{Yk})^2} , \tag{23}$$

$$\Delta X_{ij} = \left| \Delta x_{ij} \cos\theta_k + \Delta y_{ij} \sin\theta_k \right| , \tag{24}$$

$$\Delta Y_{ij} = \left| \Delta y_{ij} \cos\theta_k + \Delta x_{ij} \sin\theta_k \right| , \tag{25}$$

where Δx and Δy are the east-west and north-south distances between the i and j positions, θ_k is the orientation of the major axis of correlation in province k, and R_{X_k} and R_{Y_k} are the correlation length scales along the major and minor axes, respectively.

In cases where the covariance is required between positions in two different provinces, the covariance is computed as the product of the square roots of the covariance computed, first using parameters of one province and then the parameters of the second province.

The optimum interpolation results in a three-dimensional grid of temperature covering the domain of the grid of the circulation model. Since salinity is not available from this analysis, dynamic heights at the surface are computed from the same relationships between dynamic height and temperature—Eqs. (15) and (16)—used in modeling the structure of rings. The root-mean-square error in dynamic height computed by this method is about 0.065 dyn m. Figure 7 shows an example of the resulting dynamic height field prepared for 15 April 1987.

3.2 Conversion of Dynamic Height to Free-Surface Anomaly

The three-dimensional, thermal optimum interpolation procedure described in the previous section yields a map of dynamic height relative to some reference level. However, the analogous circulation model variables are free-surface anomaly and the depth of the interface between the two layers. Further, the spatial gradient across the Gulf Stream front implies a current that, in the ocean, will decrease with depth. The circulation model, however, will use this same current throughout the entire upper layer, which is nominally about 1000 m thick. Both factors imply a correction factor (probably spatially varying), which should be applied to the height fields created by OTIS before they are used in the circulation model.

For the sake of simplicity at this stage of development, a constant scale factor was used. Two model forecasts were made with a factor of 1.0 (that is, no correction), and the eddy shedding and meandering clearly occurred too fast. Additional forecasts were performed using scale factors of 0.666, 0.5 and 0.333 applied to four of the initial states. The effect of using scale factors smaller than 1.0 is to reduce the growth rate of meanders and the rate of ring formation. The scale factor was chosen by comparing the verification state to the *daily* output of the numerical forecasts.

As shown in Fig. 11, the larger scale factors result in the error between the forecast position of the stream and the verification state reaching a minimum earlier than it should. The model forecast at +7 days should be the best match to the verification state at +7 days. The figure shows the best match (the minimum point of the curves) occurring somewhere around the 0.4 factor. This figure includes data obtained after the original CIMREP evaluation. Using only those data, the best forecast appeared to occur with a factor of 0.333, which was used uniformly in this report. Additional studies are contemplated to further refine the transformation required between OTIS height fields and circulation model free-surface anomaly fields and to place it on firmer theoretical ground.

The way that adjusting this scale factor affects the growth rate of meanders is shown in Fig. 12. Each curve represents the persistence error between the initial state and the daily forecast out to 2 weeks, showing how the stream axis evolves away from its initial position. Each point represents an average of all 8 forecast periods. For comparison, an estimate of the *true* meander growth rate is also shown. This estimate was inferred from a year of NEOC weekly Gulf Stream axis locations by a technique described in Sec. 4.3.

Note that although the overall scale factor of 0.333 provided the best forecast skill, it actually yielded a meander growth rate that is too small. These seemingly inconsistent conclusions are an artifact of our choice of a domain-wide, constant scale factor for converting from OTIS dynamic height to OCEANS' free-surface anomaly. The absolute offset method of measuring errors is strongly affected by individual events: that is, the shedding of eddies; as such it tends to be weighted in the direction of being a measure of event prediction skill. Tuning the scale factor to minimize the offset error yields an improved forecast of events, but at the cost of a degraded forecast for the normal meanders. More technically correct methods of converting between dynamic height and free-surface anomaly are being pursued in DART. Some of these artifacts will probably be reduced and the forecast skill of the system improved even further.

3.3 Circulation Model

The basic circulation model used in this study is documented in Hurlburt and Thompson (1980) and Wallcraft (1991). Applications of the model are described in Thompson and Hurlburt (1982),

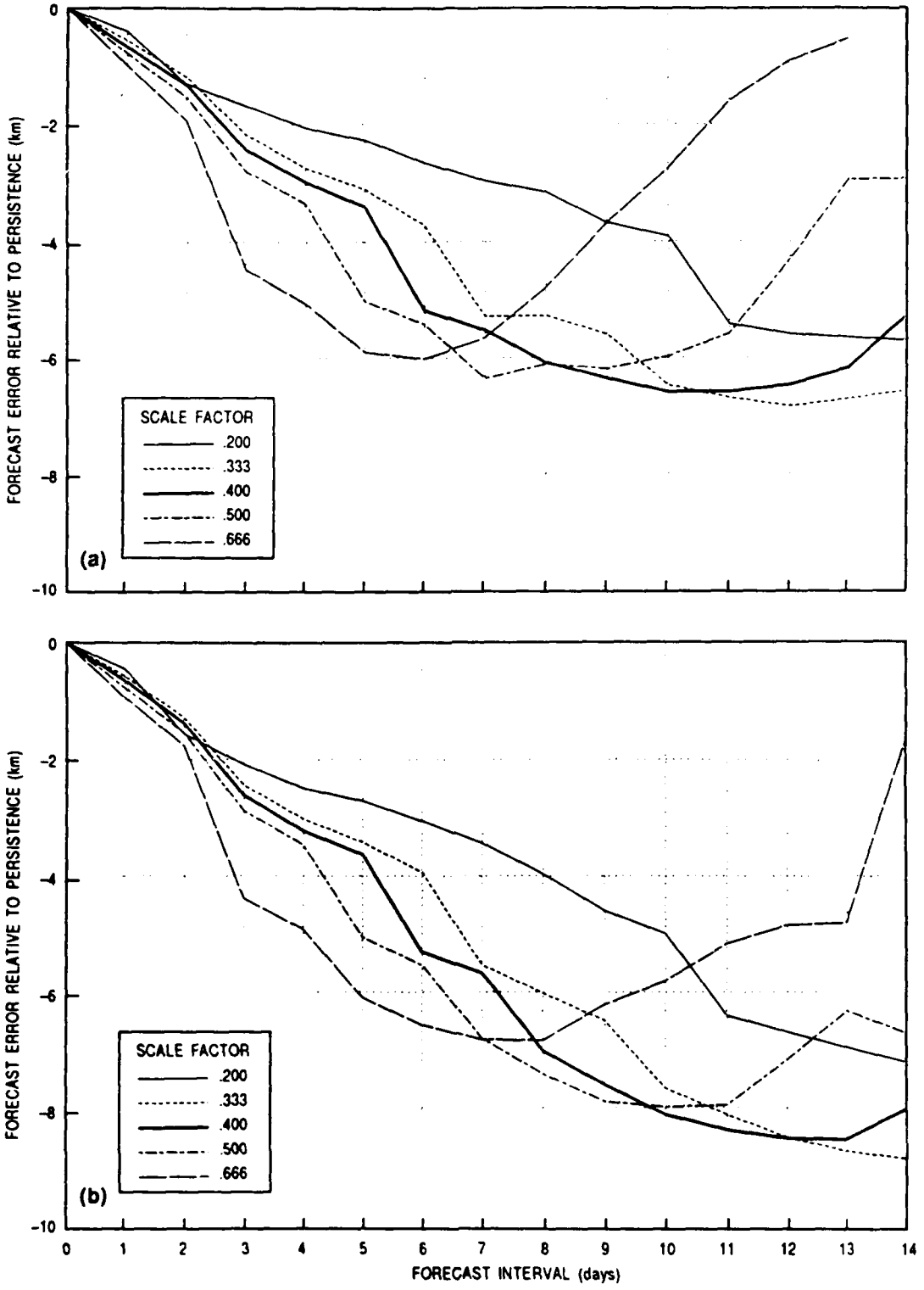

Fig. 11 — (a) Day-by-day comparison of the forecast to the verification state at +1 week, showing how the evolution of the stream axis (as measured by the average absolute offset error) changes with the choice of scale factor used to convert from dynamic height to free-surface anomaly. (b) Same, but for 2-week forecasts and verifications.

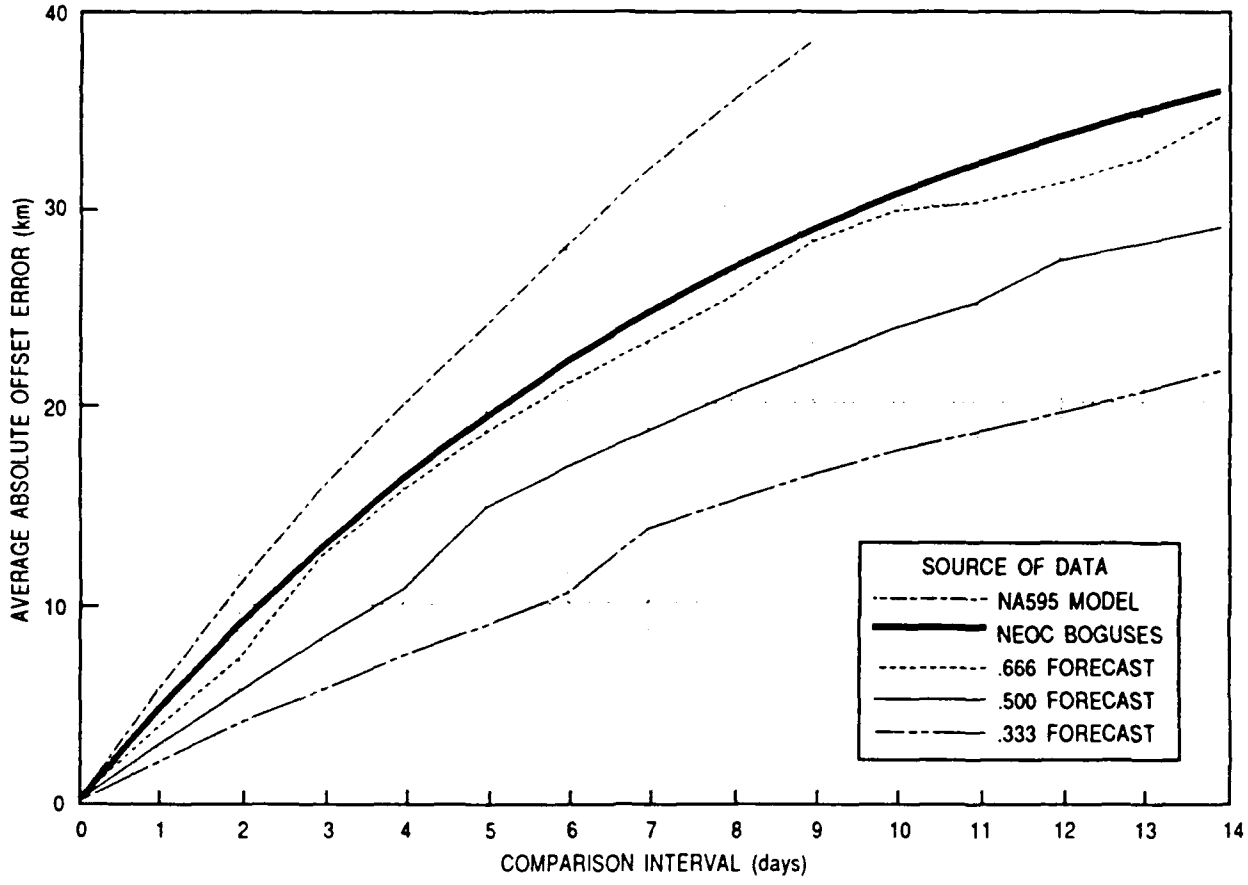

Fig. 12 — The effect on the meander growth rate of changes in the scale factor used to relate dynamic height to the free-surface anomaly. For comparison, the meander growth rate estimated from a year of NEOC boguses is also plotted. See Sec. 3.2 for a discussion.

Hurlburt and Thompson (1984), and Thompson and Schmitz (1989). It is an *n*-layer, primitive equation model covering the region from 78°W to 45°W and from 30°N to 45°N (roughly from Cape Hatteras to the Grand Banks). It includes large-amplitude bottom topography (Fig. 13). The model domain was chosen so that the variability in the location of the Gulf Stream entrance into the domain would be small. Figure 14 shows a plot of weekly axis locations taken from a year of NEOC boguses. The outlined box in the plot represents the circulation model domain. The relatively small variability in the position of the Gulf Stream axis where it enters the model domain permits the inflow to be specified at a fixed location south of Cape Hatteras. The version of the model used in this evaluation included two layers, with a deep western boundary current (Thompson and Schmitz 1989) supplied by an inflow port in the northeastern part of the lower layer. The model used in these experiments is on a spherical grid with a resolution of 1/6° in longitude and 1/8° in latitude, which represents a spatial sampling of approximately 14 km in each direction at the center of the grid.

Since layer thickness is included among the model variables, fluctuations of the pycnocline can be modeled by changes in the depth of the interface between the upper and lower layers. The layer

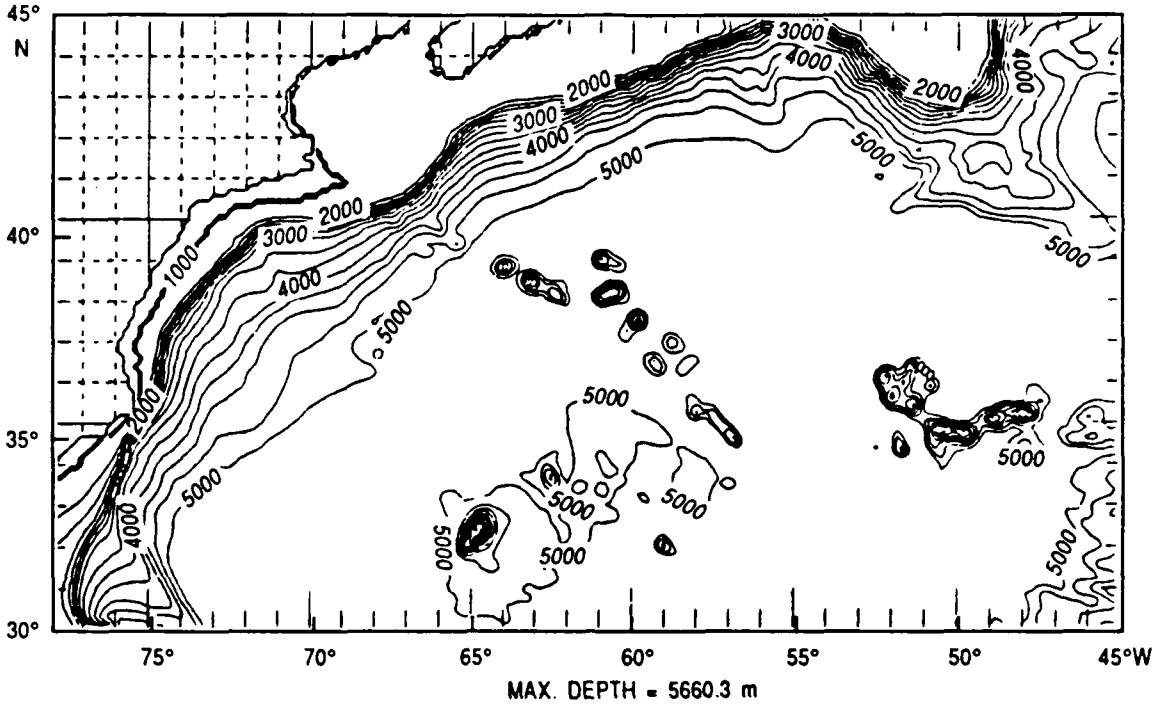

Fig. 13 — Seafloor topography used in the NRL North Atlantic regional circulation model.

thickness permits a more efficient representation of the dominant dynamical modes in the domain than is possible with a model that uses fixed thickness *levels*. This was deemed of particular importance in these experiments, due both to the manner in which we initialize the lower layer (described in detail in the next section) and the number of experiments contemplated.

3.4 Statistical Inference of Subthermocline Information

Information on the subthermocline is extremely valuable in forecasts based on numerical simulations of the Gulf of Mexico (Grant and Hurlburt 1985; Hurlburt 1987) and the Gulf Stream (Fox et al. 1988; Hurlburt et al. 1990).

In the NRL two-layer ocean circulation model, the sea surface height anomaly (η) and the pycnocline depth anomaly (h'_1) are related to the upper and lower layer density-normalized pressure anomalies (p_1 and p_2) by

$$\eta = p_1/g \tag{26}$$

$$h'_1 = (p_1 - p_2)/g' \tag{27}$$

where g is the acceleration due to gravity, $g' = g\Delta\rho/\rho$ is the reduced gravity resulting from stratification, and ρ is the density of seawater.

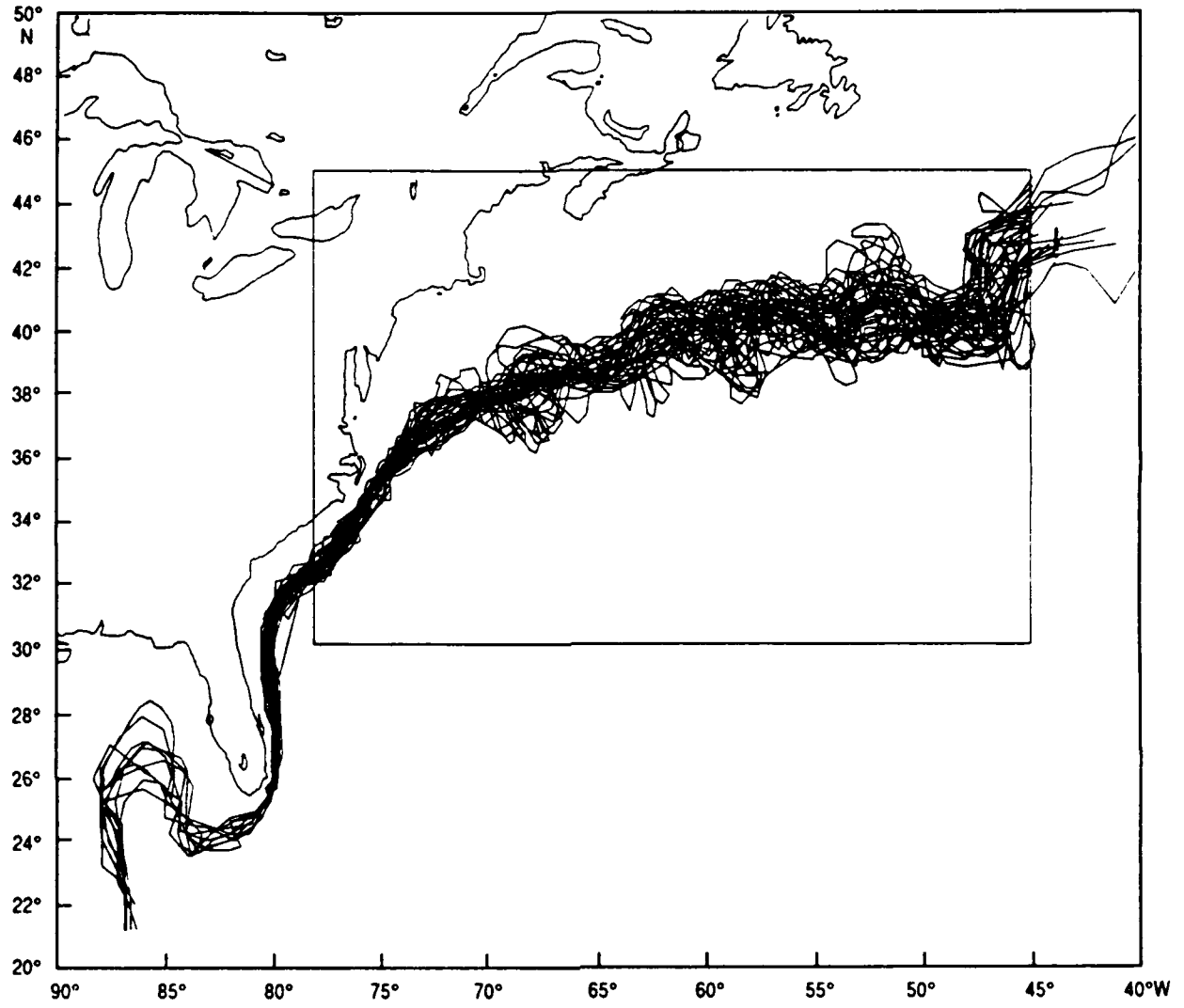

Fig. 14 — Variability in the location of the Gulf Stream frontal axis displayed by plotting a year of NEOC weekly front locations. The inner box represents the domain of the circulation model.

Long model simulations are used to derive statistical relationships between the subthermocline pressure at any given gridpoint in the model and the surface pressure at an array of gridpoints. In the present experiments, the model was spun up from rest for 17 model years, at which time both layers reached statistical equilibrium as measured by the potential and kinetic energies.

Five years of monthly fields were then used to derive EOF regression coefficients. Parameters that control this derivation are chosen to maximize the skill in estimating the lower layer pressure in an *independent* dataset. That is, coefficients are derived from one run of the model and are used to estimate the lower layer in an independent run. In the cases of the Gulf of Mexico (Hurlburt et al. 1990) and the Gulf Stream (Fox et al. 1988), the lower layer pressure anomaly from the model simulations can be accurately estimated by such techniques. Figure 15 shows an example of using these coefficients to estimate the lower layer pressure for the Gulf Stream. Note that while the

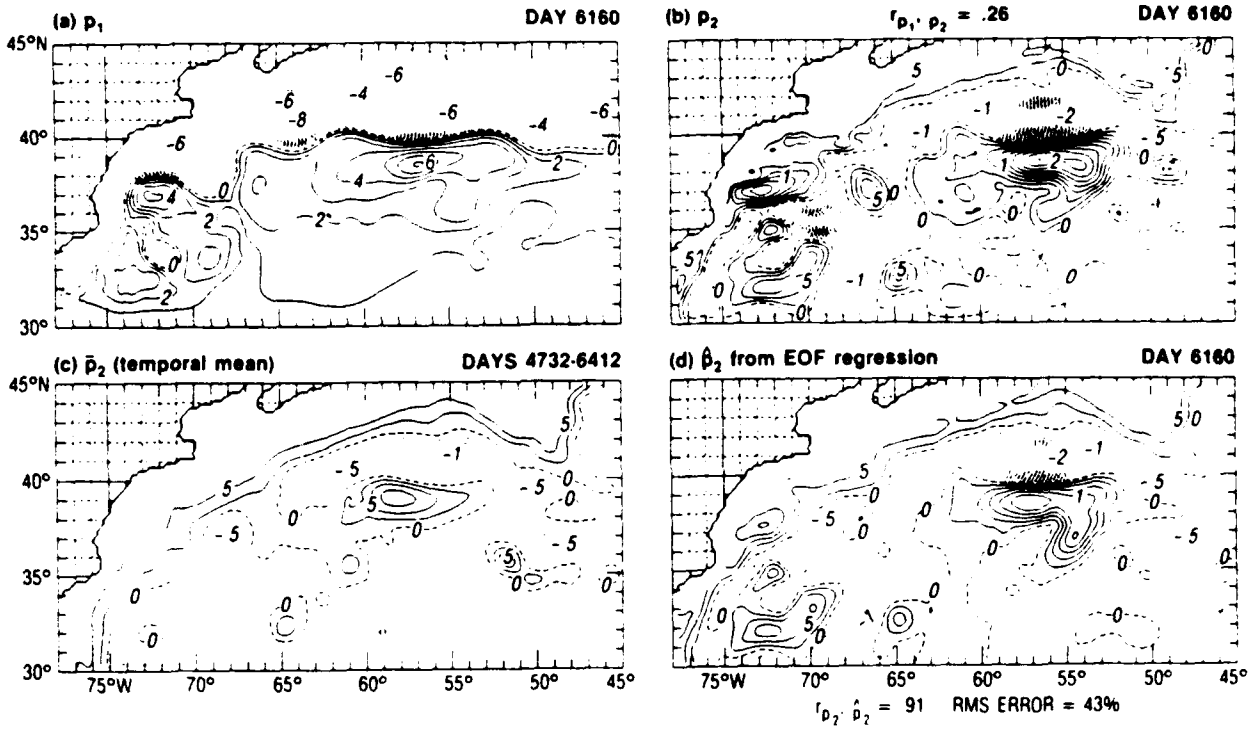

Fig. 15 — Plots of (a) upper layer pressure (p_1), (b) true layer pressure (p_2), (c) $\bar{p}_2$, a 4.5-year climatology of p_2, and (d) $\hat{p}_2$, the statistically inferred p_2 for a Gulf Stream simulation. Although the correlation between p_1 and p_2 is only 0.26, the correlation between true and inferred p_2 is 0.91. From Hurlburt et al. (1990).

pattern correlation between p_1 and p_2 on this model day is only 0.26, the correlation exceeds 0.9 between p_2 and the estimate of p_2 computed from p_1 by the statistical inference technique.

The coefficients derived from these lengthy model simulations are applied to the surface height fields produced by the thermal analysis (OTIS nowcast) to provide an estimate of the lower layer pressure field and, thus, the pycnocline depth anomaly for each of the forecast dates. The absolute accuracy of this estimate of the pycnocline depth has not been quantified, but for the purposes of initializing the circulation model, it represents lower layer information which is *dynamically consistent* with the upper layer information provided by OTIS.

Note that the statistical inference technique uses the upper and lower layer *pressure anomalies* rather than the upper and lower layer *thicknesses* because the layer thickness deviations are highly correlated, whereas at any single gridpoint the upper and lower layer pressure values are very poorly correlated. By estimating the lower layer pressure anomaly, we maximize the amount of "new" information that can be extracted from the statistical inference technique.

The contribution to the forecast skill of the DART OCEANS/GS 1.0 made by the statistical inference of p_2 was measured in comparison to two alternate methods of initializing the lower layer pressure anomaly. The first, referred to as the reduced gravity initialization, assumes no information about the lower layer and simply sets p_2 to zero in the initial state. Long model simulations indicate that the root-mean-square level of p_2 should be about 12% of the level of p_1; thus, using this

method of initialization forces the model to enter a spin-up phase to bring the lower layer to statistical equilibrium. A second method to initialize p_2 is to use the model's own climatology. Earlier experiments using numerical simulations of the Gulf of Mexico (Hurlburt 1986) and the Gulf Stream (Fox et al. 1988) indicated that the forecast skill of the model improved as information was added to the system. That is, forecasts in which the reduced gravity approximation was used to initialize p_2 showed the least skill; forecasts in which the model climatology was used to initialize p_2 showed increased skill; and forecasts in which the statistical inference technique was used showed the greatest skill.

Forecasts were made using the present reference datasets and using the three alternatives for defining the lower layer pressure field in the initial states. Table 6 summarizes the results of these experiments. With only a single exception, the statistical inference technique provided the best forecasts in all three subregions and in the overall evaluation domain for both 1- and 2-week periods. It is interesting to note that for these "real data" forecasts, the reduced gravity initialization was slightly better than using the model's own climatology for initializing p_2. Also, it should be noted that all three methods provided forecasts that were better than persistence.

Table 6 — Summary of Forecasting All the Standard Reference States But Varying the Method of Initializing the Lower Layer Pressure Anomaly

Impact of p_2 Initialization Method on Forecast Error (km)				
Results at +1 Week				
Domain	Persistence	p_2 Initialization Method		
		Red Grav.	Model Clim.	Stat. Inf.
West	24.2	23.0	22.2	18.0
Center	34.7	33.3	33.9	30.0
East	38.2	34.3	39.2	35.0
Overall	32.4	30.1	31.6	27.4
Results at +2 Week				
Domain	Persistence	p_2 Initialization Method		
		Red Grav.	Model Clim.	Stat. Inf.
West	29.3	27.7	27.4	25.9
Center	53.6	46.1	49.4	40.9
East	45.0	40.9	45.0	37.8
Overall	42.6	38.0	40.3	34.7

3.5 Geostrophic Velocity Initialization

The remaining model variables, the u and v components of velocity for each of the two layers are computed geostrophically. For example, in the upper layer:

$$\hat{k} \times f\bar{v}_g = -g\nabla\eta \,, \tag{28}$$

where f is the Coriolis parameter ($f = 2\omega \sin\theta$, where ω is the angular velocity of the earth's rotation and θ is the latitude, $\hat{k}$ is a unit vertical vector, and $\bar{v}_g$ is the geostrophic component of the current.

To examine the impact on the forecast skill of doing geostrophic velocity initialization, a series of forecasts were made based on fields taken from a long model simulation of the Gulf Stream. In each case, the model runs were compared to similar runs in which the velocities in the initial states were replaced by those computed geostrophically from the "true" p_1 and p_2 fields. The error in the forecasts was measured using the same absolute axis offset measure as was used to evaluate the forecasts that were based on actual data. Figure 16 shows a summary of the average forecast error

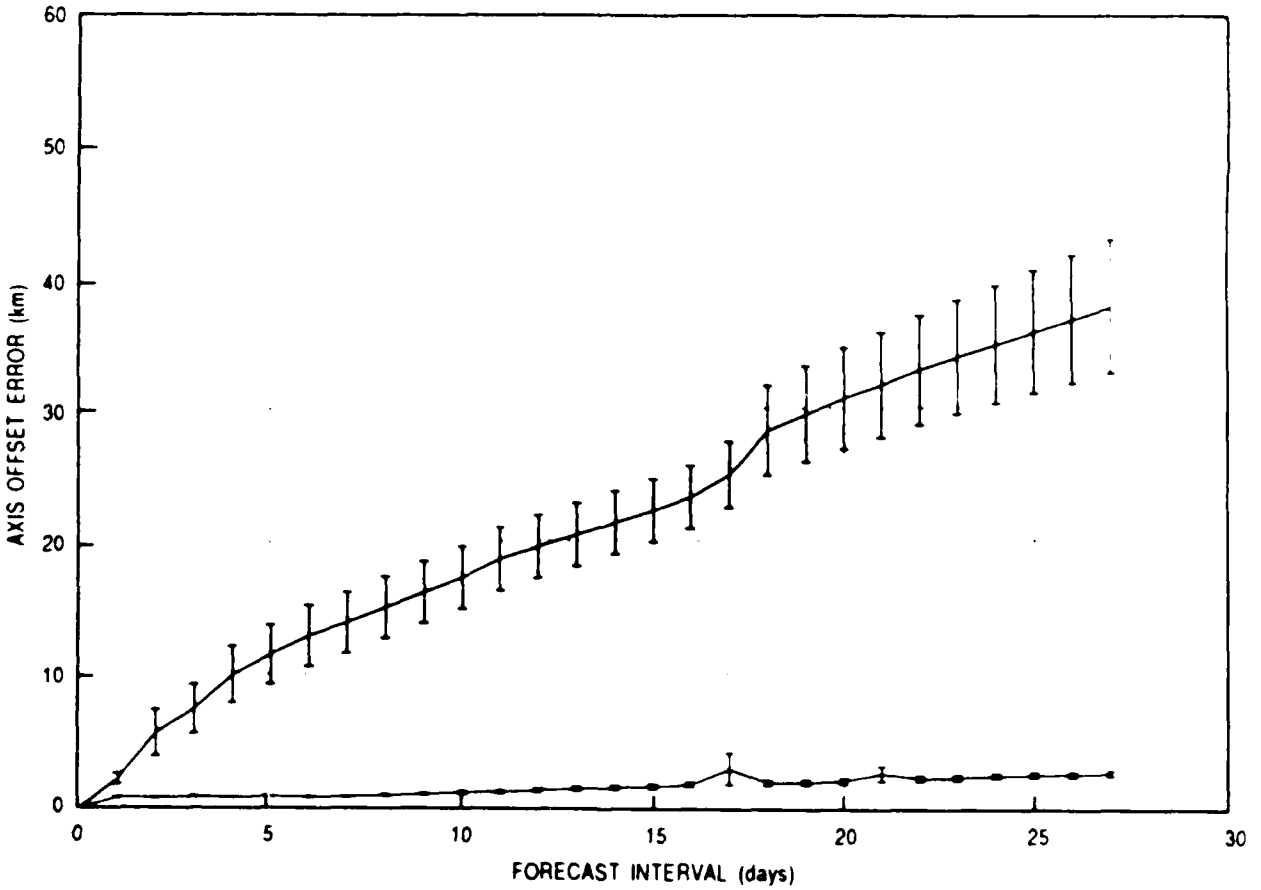

Fig. 16 — Impact on the forecast skill of doing only geostrophic velocity initialization, using model data. Top curve is average persistence error and bottom curve is average forecast error for 19 cases. Bars represent the standard error of the mean.

and the average persistence error for intervals up to 28 days long, based on 18 independent experiments. Typically, the axis has an offset error of only 2 or 3 km at +2 weeks that can be attributed to the lack of ageostrophic information in the initial state. Note that most of this error is already apparent at +1 day, with only a slight growth thereafter. It is concluded that the lack of ageostrophic information in the initial state does not contribute significantly to forecast errors.

3.6 Gravity Wave Filter

Despite the statistical inference of the lower layer and the geostrophic velocity initialization, some dynamic imbalances will inevitably exist in the initial state. One advantage of the primitive equation model approach is that such imbalances will be converted to short-period gravity waves, which can easily be removed by selective filtering. For the particular domain of this model, the dominant gravity wave period is approximately 8 to 9 h. These waves are attenuated in the NRL circulation model by a time-domain running average[2] with a time span of 8.5 h, which is applied once at 12 h into the forecast and again at 24 h, after which no further gravity wave filtering is done during the remaining 2 weeks.

Figure 17 illustrates an example taken from a gridpoint in the model, which is away from the stream in a relatively quiet region. The top pair of graphs displays the unfiltered time series of the sea surface height on the left and the Fourier amplitude spectrum on the right. The center pair displays the same information but after the application of the gravity wave filter just described. The bottom pair displays a plot of the filter itself and its impulse response. Since the filter is a simple running average (with only the two elements on the edges being reduced in amplitude), it will reduce the amplitude of frequencies well outside the period of the gravity waves; therefore, its application must be restricted. In the present implementation of the OCEANS, this filter is applied only twice and only during the initial 24 h of the forecast to minimize its effects.

4.0 ROBUSTNESS OF FORECASTS

In this section, we focus on the degree to which the accuracy of the forecasts depends on the accuracy of the initial and verification states. Unlike the observed sensitivity of GulfCast, the DART OCEANS/GS 1.0 will be shown to provide highly robust forecasts.

Section 4.1 poses, but does not answer, the question of what constitutes an "acceptable" level of error in an initial state and a forecast. Section 4.2 describes some of the sources of error in the initial and verification states, and Sec. 4.3 describes a method used to estimate the gross magnitude of this error. Section 4.4 summarizes the results of a large number of sensitivity studies, wherein a series of initial states are perturbed using an error model designed to simulate the positional errors described in Sec. 4.3. Correlated positional errors of an amplitude around 15 km are induced, and the effect of these errors on the forecast are quantified. We will show that, in many cases, the 1- and 2-week forecasts are actually *insensitive* to errors of even this magnitude. Eddy shedding and meander development and collapse are often relatively unaffected by these errors. Some forecasts, however, did exhibit significant sensitivity to changes in the initial state.

---

[2] The "running average" filter was used because it is already implemented in the NRL PE circulation model. Filters that attach the gravity wave period more selectively have been designed and will be accepted by the newer model codes.

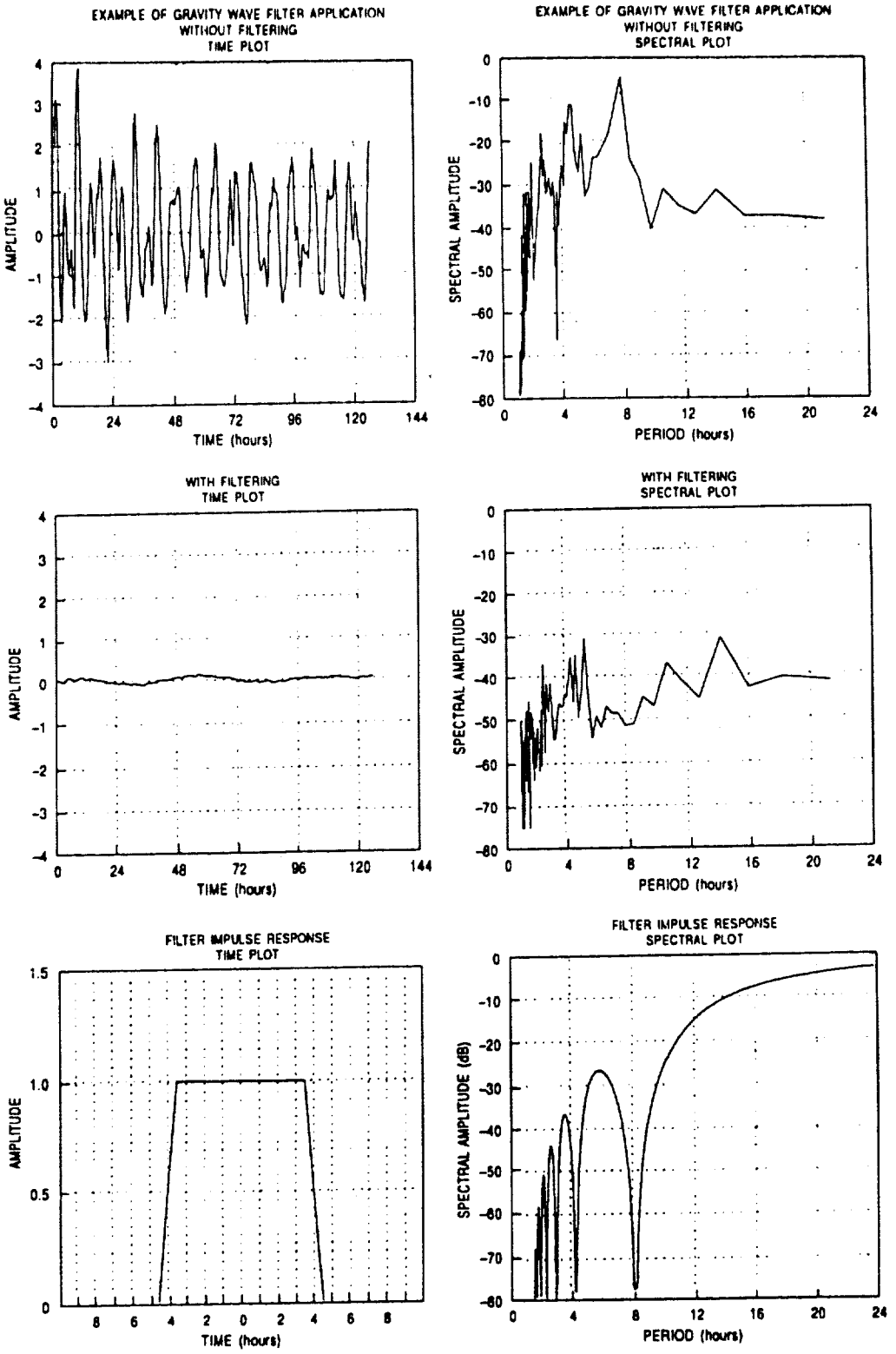

Fig. 17 — Example of the application of the filter used to remove gravity waves from the forecast. See the text in Sec. 3.6 for details.

Finally, Sec. 4.3 suggests a method to estimate the errors in any particular week's forecast, based on the experience gained in these sensitivity studies. The suggestion is based on running a series of Monte Carlo experiments to determine the regions of the Gulf Stream in which the evolution is most sensitive to small changes in the initial front and eddy maps. This information could be useful in highlighting the regions of the stream where the most accurate information is required and would thereby permit a more efficient expenditure of such resources as AXBTs and analyst working hours.

4.1 Defining the Acceptable Error Level

The determination of what constitutes an acceptable level of error in the forecast is beyond the scope of this report. For some uses, errors of around 15 km might be acceptable, but errors of even a few kilometers are unacceptable for others.

Results presented in this report show that the DART system, even in its preliminary form, provides a forecast that is better than persistence. Comparing the location of the axis in the forecast and the verification states showed that the error at 1 week was approximately 25 km. This error has a number of sources, including the error in the verification state (which will be shown to be around 12 km) and the error in the initial state (another 12 km), which will cause the model to evolve incorrectly. A large portion of the forecast axis position error (about 25 km) measured in the evaluations is therefore likely due either directly or indirectly to errors in the initial and verification states themselves. The number of in situ measurements required to define the initial position of the fronts and eddies to (say) 1-km accuracy would be astronomical. Therefore, as long as the system remains essentially a "cold start" initialization, the types of errors we see in the initial state front and eddy maps (and therefore in the forecasts as well) are unlikely to be significantly reduced.

The next DART nowcast/forecast system (DART OCEANS/GS 2.0, diagrammed in Fig. 1) will include a method of updating based on the objective assimilation of new data (satellite altimetry and other types). In this manner, the circulation model will act as an intelligent interpolator, spreading out the sparse information available from satellites and ships. The error level in the forecast will certainly be decreased as updating methods are included in future upgrades to the DART system.

4.2 Effects of Frontal Position Errors

One of the difficulties in evaluating the forecast skill of nowcast/forecast systems is deciding which errors are due to the initial and verification front and eddy maps and which are due to other factors. Errors in the position of the front might not only be amplified by the subsequent forecast but also could lead to large errors in the evaluation even when the forecast was performing well. An accurate initial state and accurate forecast would be measured against a verification state that might be erroneous. An attempt has been made to quantify both the accuracy of the initial states and the effect of typical inaccuracies on the forecast skill.

The position of the Gulf Stream front contains errors from several sources:

- The data used to define the position of the front (and eddies) is generally not synoptic. Satellite infrared photographs spread over several days might be used to find days with the least cloud cover.

• In some regions, there may be no data at all to define the front, in which case the "best guess" of the operator must be used. In some cases, persistence is used when new data are not available.

• Drawing the front can be a somewhat artistic practice, with the operator being given disjointed and perhaps inconsistent pieces of a front. These pieces must be combined into a complete picture. A manual "best fit" line is drawn through the data to provide the best educated guess as to the front position. Such programs as PATHFINDER (Horton 1989) can help to automate this process but still yield a composite front that might not necessarily be accurate.

• The process of digitizing the front from a hand-drawn map also introduces errors. Our own experience is that a root-mean-square error of as much as 8 km is introduced, even when the same operator digitizes the location more than once.

These factors act together to produce a front that is grossly correct, but one that almost certainly has positional errors associated with it.

4.3 Estimating Frontal Position Errors

Knowing that the forecast error will be due partially to the natural meander growth of the Gulf Stream, and partially to errors in defining the location of the axis, it becomes necessary to estimate the true meander growth rate (that is, how fast the error grows if we assume persistence as the forecast).

The special cases developed for the CIMREP evaluation of GulfCast were produced using far more data than is available in the operational products. These data are thus not representative of the sorts of data quality routinely available. The CIMREP standard cases had the advantages of large numbers of special AXBT surveys, more infrared data than will be routinely available and considerably more working hours of expert analysis. This section focuses on the robustness of operational forecasts, which must be initialized and verified without these advantages.

The position of the Gulf Stream axis for 1 year was extracted from standard NEOC front and eddy maps. This procedure provided 52 weekly estimates of the position as it appears in normal, operational products. The offset error was computed for delays of 1 to 6 weeks using all possible causal permutations. For example, there were 51 one-week error measures, 50 two-week errors, and so on. Figure 18, a summary of these results, clearly shows an error measure that starts low and fairly rapidly asymptotes. This plot represents data from only the central subdomain, but each of the other subdomains displays similar behavior.

Each NEOC bogus has associated positional errors. The data points presented in Fig. 18 thus represent three components: errors in the two fields being compared, plus the natural meander growth over that period. Simply extrapolating the data points in the figure shows that the error will not vanish as the time delay goes to zero.

To define the initial meander growth rate, the axis location was extracted from a long circulation model simulation run. The meander growth (as measured by the growth of persistence error) was estimated over much longer periods as well. Figure 19 shows the results of this calculation using data from the central subregion. The data are represented by the open circle symbols. Figure 19a shows the evolution for comparison intervals up to 6 months, and Fig. 19b focuses on the meander growth for intervals out to 4 weeks. The stream axis in the model evolves away from persistence

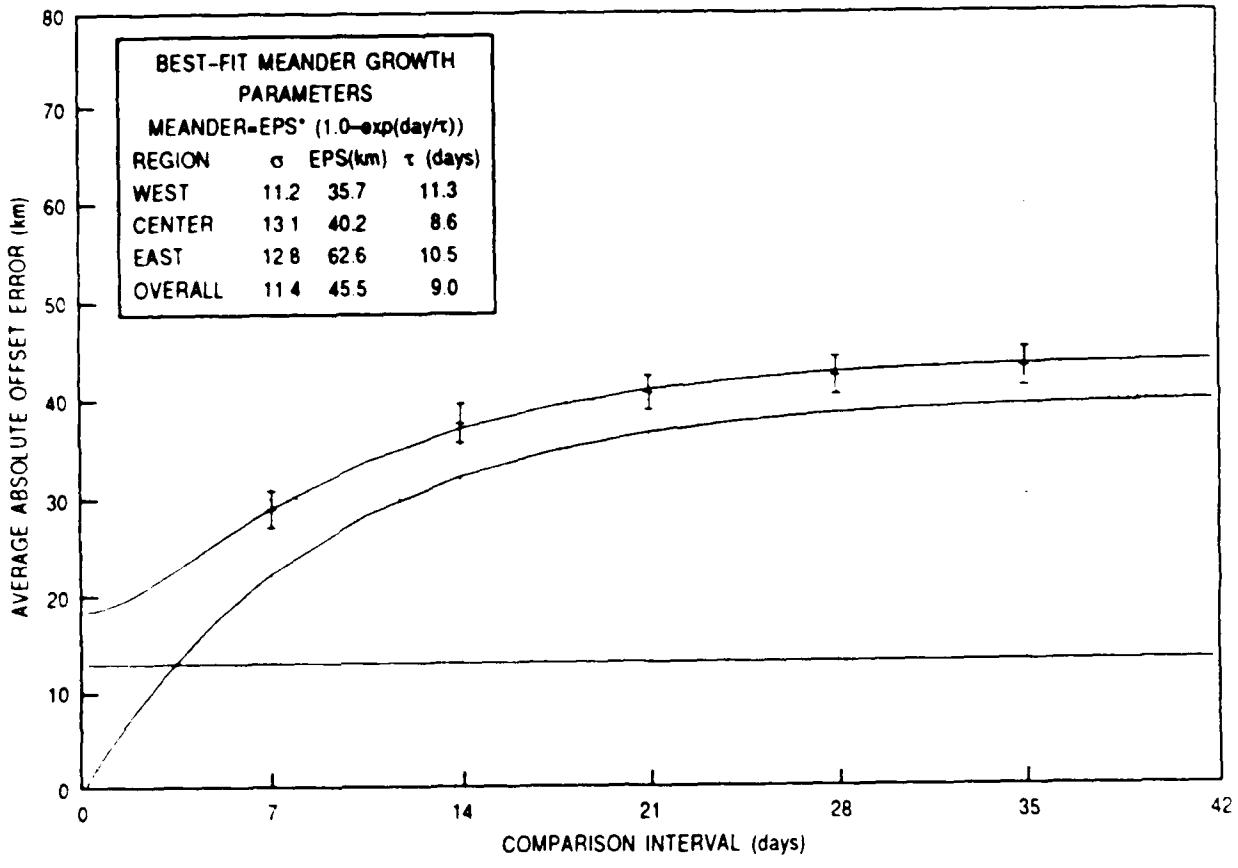

Fig. 18 — Components of persistence error computed from a year of NEOC boguses (central region). Data is represented by the symbols with error bars. Lowermost flat line is the computed average positional error in each bogus map. Middle curve is the meander growth itself, and top curve is the meander growth plus the initial and verification state errors.

toward an asymptotic limit of around 100 km. Note the shape with which this evolution proceeds. Rather than being a classic exponential shape with a "doubling time," it begins approximately as a linear growth and becomes asymptotic after a few weeks. Each of the various subdomains analyzed showed similar behavior, varying only in the asymptotic level reached and the speed with which that limit was achieved. This analysis indicates that the amplitude and time scale of meandering depends on location. Recalling Fig. 12, we see that the short-term forecasts also display meander growth curves, which begin approximately linear and show a tendency to asymptote rather than to grow exponentially. These observations will be used in the next section to define a meander growth model, which will then be applied to a large data set of actual Gulf Stream axis locations to infer approximate values of meander growth rates and errors in the axis locations.

4.3.1 The Error Model.

From the previous section, we saw that the meander growth manifests itself in the offset error measure as a term that is initially linear, but one that asymptotes after several days or weeks. A simple model fits these observations and the appearance of these curves:

$$\epsilon(t) = \epsilon_0(1 - e^{-t/\tau}) \tag{29}$$

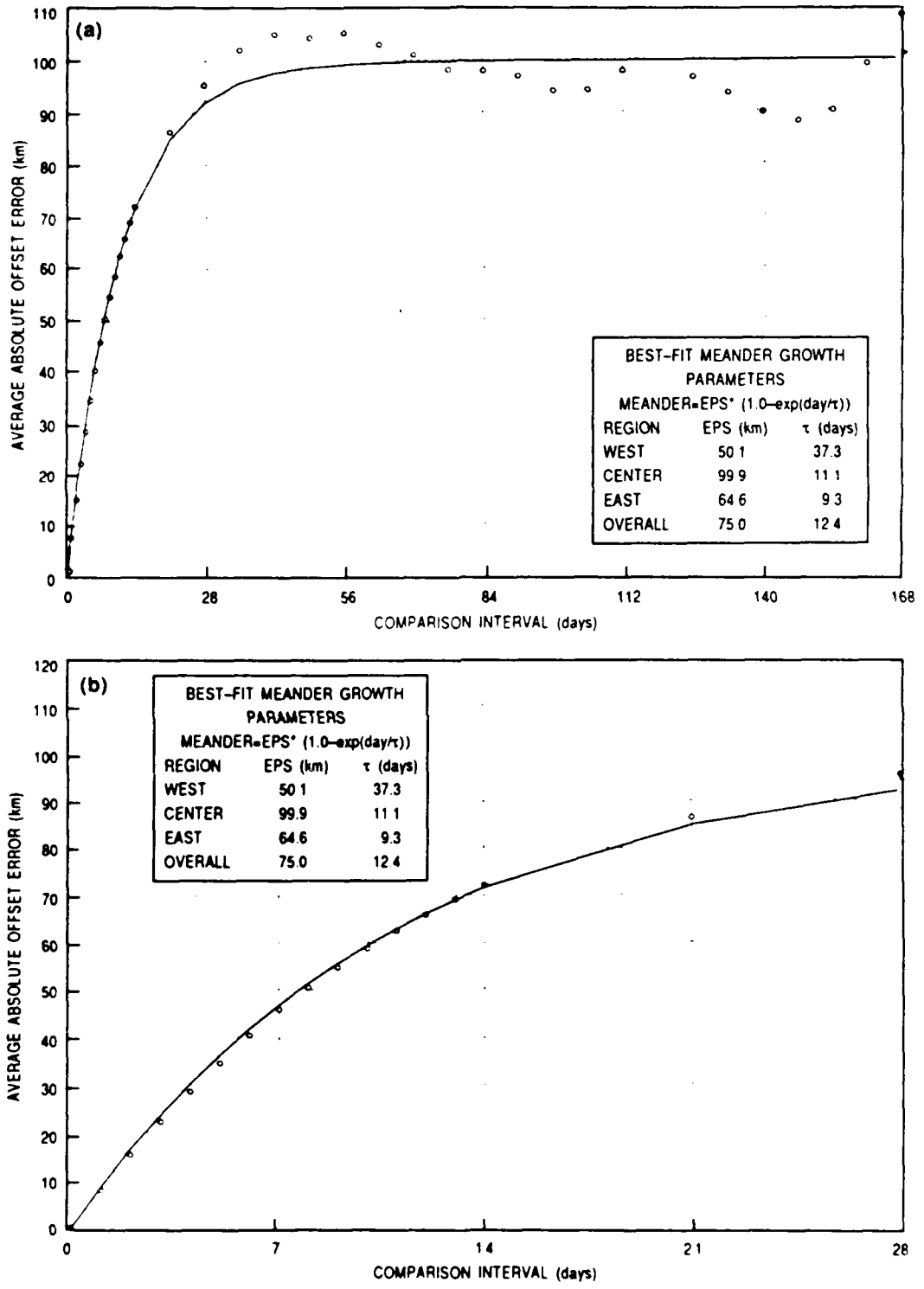

Fig. 19 — (a) Validation of the meander growth model by applying it to a year of axes extracted from a circulation model simulation. Circles represent the mean persistence error for comparison intervals ranging from 1 day to 6 months. (b) Same as in (a), but focusing on the first 28 days.

where ϵ_0 will be the error between two front locations so far separated in time as to be decorrelated, and τ describes the time scale of the evolution. It can be thought of as a decorrelation time scale for the evolution of the front. At day τ, the error has already reached 63% of its maximum value. By day 2τ the error has reached 86% of its maximum, and the two fields can probably be considered uncorrelated. This meander growth model was applied to the data in Fig. 19, and the results of the best fit are plotted using a solid line. The parameters of the fit for this subregion, as well as the others, are shown in the box.

If we had a series of absolutely accurate positions of the true Gulf Stream front, we could directly compute the parameters in Eq. (29), but we know that our boguses also include some random positional error (as described in Sec. 4.2). Assuming that this error is not correlated with the position of the front in later weeks, we can model the error between any two fields separated by a time t as

$$e^2(t) = \sigma_1^2 + \sigma_2^2 + \epsilon^2(t). \tag{30}$$

This computation says that the error between the two fields will be composed of three independent parts: the positional error (noise) in the first field, the positional error in the second field, and a term resulting from the evolution of the front between the two times. If we assume that the error in any given initial state is about the same, the total error can be modeled simply as

$$e^2(t) = 2\sigma^2 + \epsilon^2(t). \tag{31}$$

4.3.2 Results

One full year of weekly NEOC boguses was used to compute the errors between fields separated by 1 week, 2 weeks . . . 6 weeks. The error model in Eq. (31) was fit to these data and used to compute estimates for the typical errors in the initial states (σ), the decorrelation time scale for the true evolution of the stream (τ), and the error between decorrelated fields (ϵ_0).

Using procedures available in the SAS statistical analysis package (SAS Institute 1988), the individual weekly error estimates were reduced to average errors for delays of 1 through 6 weeks. A nonlinear least-squares procedure was used to directly fit these data to the model's combining error and evolution as described above. The analysis was done for the entire region, as well as the western, central and eastern regions defined by the CIMREP evaluation.

Table 7 contains the "raw" data used in the fitting process. It displays the persistence error for delays between 1 week and 6 weeks. Although axis positions from 52 bogus maps were available, the operators occasionally had no information in a particular region and simply used persistence from the previous week in their estimate of the new bogus. The single week-to-week error measurements were edited to discard measurements that were too small.

Note that in the eastern region, the persistence estimate appears to be much worse for the evolution of the stream than in the other two regions. This result might be due either to the actual evolution of the stream or to larger errors in the boguses. The following analysis will show that the errors in the initial states are essentially the same in all three regions, and that the large errors in persistence are therefore due to the actual evolution of the stream being more "vigorous" in that region, a fact confirmed by Fig. 14.

Table 7 — Average Persistence Errors for Various Delays, Based on a Year of NEOC Boguses

Persistence Errors (km) for Various Delays						
Region	Delay (weeks)					
	1	2	3	4	5	6
West	22.7	30.2	33.7	35.9	37.8	38.3
Center	29.0	37.8	40.7	42.4	43.4	44.7
East	35.7	48.7	57.5	62.3	62.2	63.9
Whole	29.4	39.4	44.1	46.5	47.2	48.1

Table 8 — Results of Best Fit of Error Model to Persistence Error Growth with Time, Using 1 Year of NEOC Boguses

Meander Growth Model Parameters				
Region	σ (km)	ϵ_0 (km)	τ (days)	Residual (km)
West	11.2 ± 1.0	35.7 ± 0.5	11.3 ± 0.9	0.22
Central	13.1 ± 2.2	40.2 ± 1.2	8.6 ± 0.9	0.47
East	12.8 ± 3.9	62.6 ± 1.2	10.5 ± 1.1	1.43
Whole	11.4 ± 0.9	45.5 ± 0.3	9.0 ± 0.3	0.05

Table 8 summarizes the results of the nonlinear least-squares best fit by region. The estimate of the error in the bogus maps is fairly consistent around 12 km. The decorrelation time scale is also fairly consistent at about 10 days. In the eastern region, the axis of the Gulf Stream evidently tends to "flap around" more than in the other regions (consistent with Fig. 14); the error between decorrelated fields is about 60 km, compared to around 40 km in the western and central regions. The fits in all the regions are quite good: the residual root-mean-square error is less than 1 km, except in the eastern region, where the error in the fit is still less than 2 km.

These estimates for typical errors in the bogus maps are consistent with estimates made by the NAVOCEANO (comparing their normal method of generating the operational axis location to in situ bathythermograph information) and the GulfCast project at Harvard. In their analysis of the accuracy of the Gulf Stream frontal axis in the initial and verification states, the GulfCast group estimated that the front is generally defined to lie in a band of uncertainty with a full width of approximately 30 to 40 km. This is consistent with an average offset error of 10 to 12 km. Comparing the operationally routine Gulf Stream frontal axis location with that derived using in situ

BT information, the operational product was estimated to have an offset error of typically 10 to 15 km (C. Horton, NAVOCEANO, pers. comm).

In addition, we can compute the rate at which the stream axis initially evolves away from its initial location by looking at the small t behavior of Eq. (31). In the limit of small t, the error growth becomes: $\in(t) \approx \in_0 t/\tau$. The term $\in_0/\tau$ has the units of km/day and gives the initial growth rate of the meanders. Using the figures from Table 8, the overall rate is 5.0 km/day. Focusing on the three subdomains, the meander growth rates are 3.2 km/day in the western region, 4.7 km/day in the central region, and 6.0 km/day in the eastern region.

Figure 18, an example of how well the error and meander growth model fits the Gulf Stream, displays the original data from Table 7 for the central region, along with the best fit using the parameters from Table 8. The short vertical bars represent the mean and standard error of persistence computed from a year of NEOC boguses for delays of 1 to 6 weeks. The line passing through these vertical bars is the final composite error model (Eq. 31). The horizontal line near the bottom is the value of σ, that is, the estimate of the typical error in the boguses themselves for the central subregion. The curve that passes through the origin is the meander growth term alone, as given by Eq. (29). In the overall domain and in each of the subregions, the meander growth model fits the persistence data computed form the year of NEOC boguses similarly well.

Comparing the meander growths from the NEOC boguses (Table 8, Fig. 18) to the parameters derived from the circulation model (Fig. 19), we note that the time scale of the evolution is consistently around 10 days in both datasets. The exception is that the evolution in the western subregion of the circulation model is considerably slower. The amplitude of the meanders is somewhat larger in the model than that inferred from the NEOC data, particularly in the central subregion where the model has meanders that are twice the amplitude seen in the data. This limitation is most likely caused by modeling the Gulf Stream with only two active layers. The model developers have since moved on to three (and more) layer versions of the primitive equation circulation model that appear to behave more realistically. Increasing the number of layers obviously will require increased computer resources, but this model will still retain a considerable advantage over the multileveled circulation models.

4.4 Sensitivity to Positional Errors in Initial States

To investigate the effects on the forecasts of errors of about 12 km in the position of the Gulf Stream axis (that is, a 25-km-wide envelope of uncertainty), a set of parallel forecasts was performed based on distorted versions of the reference initial states. Rather than drawing alternate axis positions by hand for each of these initial states, a program was written to apply a positional error model to a field. This model, normally referred to as the rubber sheet (Clarke 1990) error model because of the manner in which it performs its distortion of a field, shifts the data in a random but correlated manner. The distortion is tapered along the model boundaries so that the boundary conditions (such as the inflow transport) are not changed.

Figure 20 compares several forecasts based on 11 "rubber sheet" deformations of a particular initial state. That is, 10 unique realizations of the rubber sheet model were used to provide 11 (including the undistorted case) slightly different versions of that state. Each of these versions was used in a 2-week forecast run to display the effects on the forecast of small positional errors.

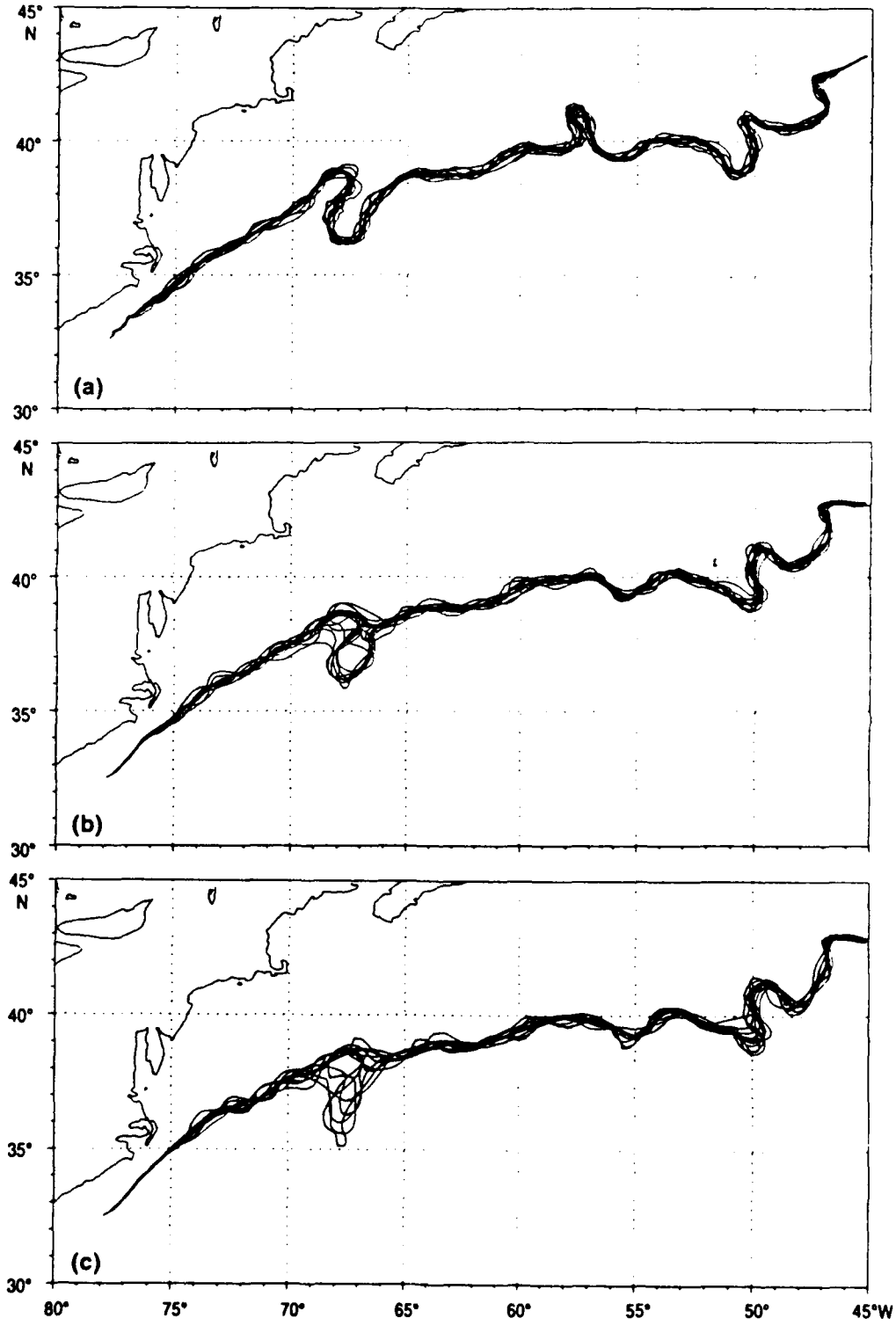

Fig. 20 — Envelope of stream axis locations from a set of forecasts based on several "rubber sheet" deformations of a particular initial state. Figure (a) shows the axis positions in the 11 initial states. Figures (b) and (c) show the axis positions after 1- and 2-week forecasts, respectively. Note the general insensitivity of the forecast to small errors in the initial states except in isolated regions.

Immediately apparent in the figure is the fact that the overall envelope that contains the Gulf Stream axis position does not appear to enlarge with time. The average absolute offset errors of 12 km in the initial states increase to only 14 km at +2 weeks. Performing these deformations on all the CIMREP datasets, we find that even eddy/stream interactions (shedding and absorption) are sometimes not significantly affected by these positional errors.

Note that for each of the initial states there was always a distorted version that yielded an improved forecast (as verified by comparison to the undistorted verification states). That is, it was always possible to create an initial state for which the axis location was within the 12-km error bounds (and therefore still a legitimate version of the data) and for which the forecast improved upon the undistorted case by at least 10 km. For each alternate initial state that yielded an improved forecast, however, there was also a version that yielded a degraded forecast. In any event, such a methodology (referred to as *hind*casting) is not relevant to the problem of *fore*casting, for which no data are ever available. Hindcasting is useful in creating improved versions of past datasets but provides no useful forecast information.

4.5 Operational Monte Carlo Forecast Scenario

The sensitivity studies suggested a method for doing operational forecasts that can provide not only an estimate of the evolution of the Gulf Stream, but also a confidence range. In one scenario, the initial state for "today" can be created, and a series of rubber sheet forecasts can be made in a Monte Carlo procedure. The results of these forecasts will immediately show where the modeled evolution of the stream is most sensitive to its initial state, and will thus highlight the regions that require special attention. In Fig. 20, for example, the forecast (based on the initial state on 06 May 1987) shows regions of eddy shedding around 67°W and 57°W, which appear to be at least somewhat sensitive to the initial state. Depending on the operational requirements for forecast skill in these locations, additional effort in improving the initial states should clearly be focused on these two areas. Numerous similar meander situations in the other initial states, however, do not show this sensitivity. A Monte Carlo approach to forecasting could thus provide a quick way of getting useful information on where to deploy resources (bathythermographs and operator care in creating the initial field).

5.0 COMPUTATIONAL REQUIREMENTS

This section addresses the computation requirements of the DART OCEANS/GS 1.0 system. In summary, it would require approximately a full day of computer time on a desktop computer, such as a Hewlett-Packard (HP) 9000/835 or a Sun SPARCstation 1, to perform a 2-week Monte Carlo forecast as outlined in this section. The same set of forecasts on the CRAY Y-MP at NAVOCEANO would require only about 0.5 h. If an operational bogus is prepared every 3 or 4 days, only a one-fourth to one-third of the time available on the HP 9000/835 would be required to perform the Monte Carlo forecasts. This technique is therefore viable even on relatively small computers.

The complete DART OCEANS/GS 1.0 was developed and evaluated on several computers, including a VAX 8800, an Alliant FX/80, an HP 9000/835, Sun workstations, and the Cray Y-MP 8/8128 at NAVOCEANO. The largest component of the system is the primitive equation circulation model, which requires approximately 8 Mbytes of main memory to be able to run without the need to swap to disk. The requirements for OTIS 2.1 are approximately half that. The system can thus be run on typical desktop computers. Except for the Y-MP, on each of these machines, approximately

50 min of central processor time is required to perform a 1-week forecast.[3] Using a single processor on the Y-MP, a 2-week forecast (including initialization) is completed in about 90 s.

All software and subroutine libraries included in the system were either developed within the Navy or are public domain (such as EISPACK and LINPACK): no vendor proprietary licenses are required. This option will facilitate the eventual movement of the system to the Tactical Environmental Satellite System, which is capable of performing the forecasts described in this report.

6.0 SUMMARY AND CONCLUSIONS

The DART project team at NRL developed and transitioned to operations the first Gulf Stream forecasting system that provides skillful estimates for locating the stream out to 2 weeks using only standard, operationally available data.

Key developments that permitted this achievement included the following.

- The development at NRL of a primitive equation ocean circulation model of the Gulf Stream region that properly represents the mesoscale features, the patterns of variability, and the evolution time scales of the region.
- The proper and optimum use of satellite-derived data (such as infrared and altimetry) to produce a field of surface dynamic height.
- The use of this surface information to initialize the multilayer circulation model in a dynamically balanced and consistent manner.

The dynamically balanced initialization of the circulation model was made possible by the development at NRL of a technique to relate surface and subsurface pressure anomalies. Residual small imbalances are removed during the first 24 h of a forecast using a short-period gravity wave filter.

Wherever possible, the DART project has attempted to use standard operational data products, such as the bogus messages created at NAVOCEANO, and computer programs, such as OTIS and OCEANS, to minimize the impact on the operational centers.

Although the system has the capability of using in situ surveys, which are expensive and resource consuming, it has shown skill in forecasting the Gulf Stream using routinely available and relatively inexpensive operational data provided by satellites. Using a Monte Carlo approach, we have shown that small errors in the estimated location of the Gulf Stream usually do not amplify into large errors in the forecast. This same approach can be used to highlight those isolated areas where additional care in the analysis of existing satellite data or the acquisition of in situ data could improve the forecast.

---

[3] In the case of the Alliant, only a single computational element is used. If a cluster size of six elements is used and the program run in parallel mode, the run time is reduced to about 15 min per forecast week.

7.0 RECOMMENDATIONS

The approach taken by NRL for developing nowcast and forecast systems has proven to be successful in the Northwest Atlantic. The applicability of the approach must now be evaluated in other regions. Initially, the other major western boundary current the Kuroshio will be examined. From there, basin and finally global scale nowcast/forecast systems will be built.

An "error budget" for the present forecast system must be developed to determine the factors that might limit further improvements. The standard operational boguses for the Gulf Stream region have positional errors of about 10 km, so that the 1-week forecast skill is expected to have a lower limit of about 15 km. Since the actual errors are higher, the source of the additional error must be in the various modules that make up the system. Each of these sources should be investigated to make certain that their contribution to the forecast skill is maximized.

8.0 ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, Navy Ocean Modeling Program, under Program Element 0602435N, project RM35G84. Mr. Robert Peloquin was the program manager. We acknowledge the outstanding work of the other members of the DART Project Team; this work led to significant advances in the state-of-the-art of ocean forecasting: Mr. Ted Bennett (NAVOCEANO), Dr. Donna Blake (formerly of the Naval Oceanographic and Atmospheric Research Laboratory), Ms. Jan Dastague (Sverdrup Technology), Mr. Mike Eckhoff (Sverdrup Technology), Dr. John Harding (NRL), Mr. Jeff Hawkins (NRL), Dr. George Heburn (NRL), Dr. Harley Hurlburt (NRL), Mr. Gary Ransford (Sverdrup Technology), Mr. Bob Rhodes (NAVOCEANO), Mr. Ole Martin Smedstad (Sverdrup Technology), the late Dr. Dana Thompson (NRL), and Alan Wallcraft (Planning Systems, Inc.).

9.0 REFERENCES

Bennett, T. J., M. R. Carnes, P. A. Phoebus, and L. M. Reidlinger, "Feature Modeling: The Incorporation of a Front and Eddy Map into Optimal Interpolation-Based Thermal Analyses," Naval Research Laboratory, Stennis Space Center, MS, NORDA Report 242, 1988.

Bennett, T. J. and P. May, "An Optimal Thermal Analysis System for the Naval Oceanography Program," Naval Research Laboratory, Stennis Space Center, MS, NORDA Report 206, 1988.

Born, G. H., J. L. Mitchell, and G. A. Heyler, "GEOSAT ERM – Mission Design," *J. Astronautical Sci.* **35**, 119–134 (1987).

Bretherton, F. P., R. E. Davis, and C. B. Fandry, "A Technique for Objective Analysis and Design of Oceanographic Experiments Applied to MODE–73," *Deep Sea Research* **23**, 559–582 (1976).

Carnes, M. R., J. Mitchell, and P. deWitt, "Synthetic Temperature Profiles Derived from GEOSAT Altimetry: Comparison with AXBT Profiles," *J. Geophys. Res.* **95**, 17979–17992 (1990).

Clancy, R. M. and P. J. Martin, "The NORDA/FLENUMOCEANCEN Thermodynamic Ocean Prediction System (TOPS): A Technical Description," Naval Research Laboratory, Stennis Space Center, MS, NORDA Technical Note 54, 1979.

Clancy, R. M. and P. J. Martin, "Synoptic Forecasting of the Oceanic Mixed Layer Using the Navy's Operational Environmental Data Base: Present Capabilities and Future Applications," *Bull. Am. Meteor. Soc.* **62**, 770–784 (1981).

Clancy, R. M. and K. D. Pollak, "A Real-Time Synoptic Ocean Thermal Analysis/Forecast System," *Prog. in Oceanography* **12**, 383–424 (1983).

Clancy, R. M., K. D. Pollack, J. D. Cummings, and P. A. Phoebus, "Technical Description of the Optimum Thermal Interpolation System (OTIS) Version 1: A Model for Oceanic Data Assimilation," Fleet Numerical Oceanography Center, Monterey, CA, Technical Note 422–86–02, 1988.

Clarke, K. C., *Analytical and Computer Cartography* (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1990).

Crow, E. L., F. A. Davis, and M. W. Maxfield, "Statistics Manual," Dover Books edition of NAVORD Report 3369 NOTS 948, 288 pp, U.S. Naval Ordnance Test Station, Bureau of Ordnance, 1960.

Cummings, J., FNOC unpublished report, 1989.

Davis, R. E., "Predictability of Sea Surface Temperatures and Sea Level Pressure Anomalies Over the North Pacific Ocean," *J. Phys. Oceanogr.* **6**, 249 (1976).

deWitt, P. W., "Modal Decomposition of the Monthly Gulf Stream/Kuroshio Temperature Fields," Naval Oceanographic Office, Stennis Space Center, MS, TR 298, 1987.

Fox, D. N., H. E. Hurlburt, J. D. Thompson, Z. R. Hallock, and G. A. Ransford, "Oceanic Data Assimilation and Prediction Using Remotely Sensed Data," Proceedings of the Pacific Congress on Marine Science and Technology (PACON 88), 1988, pp. OST2/17–23.

Gandin, L. S., Objective Analysis of Meteorological Fields, Leningrad: Gedrometeorizdat, 1963, 266 pp.

Gardner, Geraldine, Program AXERR, Harvard University, 1989.

Grant, D. P. and H. E. Hurlburt, "Statistical inference of Deep Pressure Fields from Simulated Altimeter Data," *EOS* **66**, 1324 (1985).

Horton, C. W., "An Overview of the Reliability of the Gulf Stream PATHFINDER Product," U.S. Naval Oceanographic Office, Stennis Space Center, MS, internal document, 1989.

Hurlburt, H. E. and J. D. Thompson, "A Numerical Study of Loop Current Intrusions and Eddy Shedding." *J. Phys. Oceanography* **10**, 1611–1651 (1980).

Hurlburt, H. E. and J. D. Thompson, "The Dynamics of the Loop Current and Shed Eddies in a Numerical Model of the Gulf of Mexico," in *Hydrodynamics of Semi-Enclosed Seas*, J. C. J. Nihoul, ed. (Elsevier, Amsterdam, 1982).

Hurlburt, H. E. and J. D. Thompson, "Preliminary Results from a Numerical Study of the New England Seamount Chain Influence on the Gulf Stream," in *Predictability of Fluid Motions*, G. Holloway and B. J. West, eds. (American Institute of Physics, New York, 1984).

Hurlburt, H. E., "Dynamic Transfer of Simulated Altimeter Data into Subsurface Information by a Numerical Ocean Model," *J. Geophys. Res.* **91**(C2), 2372–2400 (1986).

Hurlburt, H. E., "The Ocean Prediction Problem and Its Diversity: Some Issues and Possible Solutions," in *Ocean Prediction Workshop*, 1986, C. N. K. Mooers, A. R. Robinson, and J. D. Thompson, eds., W. Simmons, tech. ed., Sponsored by the Oceanographer of the Navy and the Office of Naval Research, pp. 192–226, 1987.

Hurlburt, H. E., D. N. Fox, and E. J. Metzger, "Statistical Inference of Weakly Correlated Subthermocline Fields from Satellite Altimeter Data," *J. Geophys. Res.* **95**(C7), 11375–11409 (1990).

Joyce, T. M., "Velocity and Hydrographic Structure of a Gulf Stream Warm-Core Ring." *J. Phys. Oceanogr.* **14**, 936–947 (1984).

Kindle, J. C., "Sampling Strategies and Model Assimilation of Altimetric Data for Ocean Monitoring and Prediction," *J. Geophys. Res.* **91**, 2418–2432 (1986).

Levitus, Sydney, Climatological Atlas of the World Ocean, National Oceanic and Atmospheric Administration, Rockville, MD, 1982.

Lybanon, M. and R. Crout, "The NORDA GEOSAT Ocean Applications Program," Johns Hopkins Applied Physics Laboratory, *Technical Digest* **8**, 212–218 (1987).

Mitchell, J. L., Z. R. Hallock, and J. D. Thompson, "REX and GEOSAT: Progress in the First Year," Johns Hopkins APL *Technical Digest* **8**, 234–244 (1987).

Preisendorfer, R. W., "Principal Component Analysis in Meteorology and Oceanography," in *Developments in Atmospheric Science*, vol. 17 (New York: Elsevier Science, 1988).

Ransford, G. A., M. A. Eckhoff, and H. E. Hurlburt, "Ocean Circulation, Altimeter Data and an Inverse Problem," *EOS* **68**, 1326–1327 (1987).

Rhodes, R. C. and G. W. Heburn, "Validation of the Harvard Model in the Gulf Stream Meander and Ring Region," Naval Research Laboratory, Stennis Space Center, MS, in prep.

Rhodes, R. C. and C. W. Horton, "Navy Operational Gulf Stream Forecast System (NOGUFS), Version 1, OPCHECK Evaluation," Naval Oceanographic Office, Stennis Space Center, MS, Report NOGUFS 90–01, 1990.

Robinson, A. R., M. A. Spall, W. G. Leslie, L. J. Walstad, and D. J. McGillicuddy, "Gulfcasting: Dynamical Forecast Experiments for Gulf Stream Rings and Meanders, November 1985–June 1986," in Harvard Open Ocean Model Reports, Reports in Meteorology and Oceanography, Number 22. Harvard University, Division of Applied Sciences, Cambridge, MA, 1987.

SAS Institute, Inc., SAS/STAT User's Guide, Release 6.03 Edition, SAS Institute, Inc., Carey, NC, 1988.

Teague, W. J., M. J. Harron, and P. J. Hogan, "A Comparison between the Generalized Digital Environmental Model and Levitus Climatologies," *J. Geophys. Res.* **95**(C5), 7167–7183 (1990).

Thompson, J. D. (1986), "Altimeter Data and Geoid Error in Mesoscale Ocean Prediction: Some Results from a Primitive Equation Model," *J. Geophys. Res.* **91**, 2401–2417 (1986).

Thompson, J. D. and H. E. Hurlburt, "A Numerical Study of the Influence of the New England Seamount Chain on the Gulf Stream: Preliminary Results," in Proceedings of the Workshop on the Gulf Stream Structure and Variability, University of North Carolina, Chapel Hill, NC, J. M. Bane, Jr., ed., Office of Naval Research, Washington, DC, 1982, pp. 346–362.

Thompson, J. D. and W. J. Schmitz (1989), "A Limited-Area Model of the Gulf Stream: Design, Initial Experiments, and Model-Data Intercomparison," *J. Phys. Ocean.* **19**(6), 791–814 (1989).

Wallcraft, A. J., "The NAVY Layered Ocean Model Users Guide," Naval Research Laboratory, Stennis Space Center, MS, NOARL Report 35, 1992.